

Scientific and Technical Information Center

SEARCH REQUEST FORM

Date: _____ Requester's Full Name: _____ Examiner #: _____

Art Unit: _____ Phone (301) _____ Serial Number: _____

Results Format Preferred (circle): PAPER DISK E-MAIL

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Date: _____

Search Topic:

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known.

For Sequence Searches Only Please include all pertinent information (parent, grandchild, divisional, or issued patent numbers) along with the appropriate serial number.

FOR OFFICIAL USE ONLY

STAFF USE ONLY

Searcher: M. Smith 83278

Searcher Phone #: _____

Searcher Location: _____

Date Searcher Picked Up: 1-17-2000

Date Completed: 1-14-2000

Searcher Prep & Review Time: 15

Online Time: 15

Type of Search

2 NA Sequence (#)

1 AA Sequence (#)

Structure (#)

Bibliographic

Litigation

Fulltext

Other

Vendors and Cost

STN Dialog

Questel/Orbit Dr.Link

Lexis/Nexis Westlaw

WWW/Internet

In-house sequence systems (list)

Other (specify)

CompuGen

This Page Blank (uspto)

[illegible]

RESULT 2

W22212 ID W22212 standard; Protein: 622 AA.

AC W22212;

DT 02-NOV-1997 (first entry)

DE Rat transferrin receptor.

KW Radiation; radiomimetic; exposure; transferrin receptor;

RM red blood cell; rat.

OS Rattus norvegicus.

PN W09715830-A1.

PD 01-MAY-1997.

PF 21-OCT-1996; U16880.

PR 24-OCT-1995; US-547197.

PA (UJNY) UNIV NEW YORK STATE RES FOUND.

PI Glomski CA, Gong JK;

DR WPI: 97-239141/23.

PT Detecting exposure to radiation or radio-mimetic agents - by

PT correlation with a protein, such as transferrin receptor, present on

PT the surface of red blood cells of a mammal

PS Disclosure: Page 61-62; 83pp; English.

CC This polypeptide comprises the rat transferrin receptor (TR).

CC Claimed methods for detecting a mammal's prior exposure to

CC radiation or radiomimetic agents (Ras) involve: isolating red blood

CC cells from the mammal, and detecting the quantity of TRs on the red

CC blood cells using e.g. an antibody raised against the TR. The

CC quantity of TRs on the red blood cells is correlated with the

CC mammal's prior exposure. It is possible to assess accurately

CC cumulative lifetime exposure and to directly study the effects of

CC low doses of radiation, e.g. on cancer risk and other delayed

CC responses. The limited amount of blood necessary permits the test

CC to be conducted as frequently as desired. The test's sensitivity,

CC extending from 0.1 cGy to over 600 cGy, spans both the environmental

CC and the lethal dose ranges.

Q0 Sequence 622 AA;

```
Query Match      33.9%   Score 1424.5; DB 1; Length 622;
Best Local Similarity 46.5%; Pred. No. 9.2e-118;
Matches 304; Conservative 98; Mismatches 209; Indels 43; Gaps

QY    155 DTIRKTS---LRERVAGSGAALATQDIRRALRSKQKLDPHTWTDTTHVYGLOPPDPAHNTL 2111
```

| | | | |
|----|-----|-------------------------------------------------------------|------|
| Db | 5 | DIHQLOSONYTPREAGSÖQKÖDENÄXYIENLFHFKRSKYMREHYYKIOYKNSVQN-L | 63 |
| Oy | 212 | HWMDAGKVGSEQLPEDDPYVCYSAIGNVTGELVYAHYGRPELDLRLARGVDPGRLL | 2717 |
| Db | 64 | VTINSQNMID---EVEAPEGVAAFSKAGEVYKLVYHNFCTKKDFEL--NYSVMSGLV | 1171 |
| Oy | 272 | LYRQGVJSEAKVYMNADFGAGUQLYPERADFSQDDPRKSLSSQOAVGVHVLGTDPY | 3331 |
| Db | 118 | IVRAGKTFEAKVMAAOSFNAGVLYIMDRNF-----PVENDLOFEGHAHGTDPY | 1717 |
| Oy | 332 | TPGEFPMQOTPPVASSGGLPSIPACEISADIASRLRLKGPAPAPOMOGSLLGSYHL | 3919 |
| Db | 172 | TPGEFSFHHTQPPSSQGLPSIVQVITISRAPKLKKMBGN-CPSQNMID-SSCKLEL | 2298 |
| Oy | 332 | GPÖPRLRVNNHRTSTP INNIFFCIEGRSEPDHYVVGARDAWGPAAKSAVGATILL | 4511 |
| Db | 230 | SÖNQNVKTLVNNVAKETRLINIEFVIGYEEPDXYIVVGAORDAWGPAAKSAVGATILL | 2881 |
| Oy | 452 | ELVTFESQMSV-NQFRPRSLSTFSMGGDGSVSTEMEGYSLVHLAKVAVVYSIDNA | 5101 |
| Db | 290 | KLAQVFSQMISKDQFRBRSITTFMSWAGDGAAGPEWMEGYSLSLHLAFTYINIDKY | 3498 |
| Oy | 511 | VIGDDKFFAKKSPLLTSLIESVLAQVDSPNHSGOTLEQVFTNPMSDAEYIRPLPMDSS | 5707 |
| Db | 350 | VIGUSNFVSAVSLPYLTLMKIMQDVNHP-IDGKYLVR-----NSNMISK-IEELSIDNA | 4021 |
| Oy | 571 | AYSPFAFVGVAVFESPEDDQATPFYHTEKEDYENLHKVLOGRLPVAQAV--AOLAG | 6271 |
| Db | 403 | AFPLFVLAAGIPAVSFCECED-YPYIGTKIDYTE---ILIDKVPOLNOMVRTAAEVAG | 4571 |
| Oy | 628 | QLLRLSHDRLPLPDEFGYGVVYRHLIGNLNEFSGDKARGLTLOWYASARGDVIRAAEK | 6871 |
| Db | 458 | QFIKTLHDIELTLEYKYNKSLSPKMDLQCFADIKDMLSLQMLYSARGDYFRATSR | 5171 |
| Oy | 668 | LQÖIYSSEERDEBLTRMYNVRIMRVEFFYLQSYVSPADSFPRIHFMGRDHTLGALLD | 7471 |
| Db | 518 | LTTDEHNAEKTNREYMEINDRIKAVEYHFLSPVSPRESFRIHFMGSSHTLALVEN | 5771 |
| Oy | 748 | LRLRSNSSGTPGATSSGPOESFRRQALITTTLOGAANALSQDVWINDNF | 8011 |
| Db | 578 | LRLQKN-----ITAFETELFRNQLALATWTIGCVANALSGDVIWNIDNE | 6221 |

| RESULT | 3 | |
|--------|-------------------------------------------------------------------|-------------------------------------------------|
| ID | R55097 | |
| AC | R55097 | |
| DT | 11-JAN-1995 | (first entry) |
| DE | Prostate-specific membrane antigen. | |
| KW | Prostate-specific membrane antigen; PSM; prostate cancer; | |
| KW | transmembrane glycoprotein; imaging; targeting; tumour detection; | |
| KW | antibody detection. | |
| OS | Homo sapiens. | |
| FT | Key | Location/Qualifiers |
| FT | peptide | 63..68 |
| FT | | /note="region of high hydrophilicity 1,claim 35 |
| FT | peptide | 132..137 |
| FT | | /note="region of high hydrophilicity 2,claim 35 |
| FT | peptide | 482..487 |
| FT | | /note="region of high hydrophilicity 3,claim 35 |
| FT | domain | 23..39 |
| FT | | /note="specific membrane spanning domain" |
| PN | W09409820-A. | |
| PD | 11-MAY-1994. | |
| PF | 05-NOV-1993; | U10624. |
| PR | 05-NOV-1992; | US-973337. |
| PA | (SLOK) SLOAN KETERING | INST CANCER. |
| PI | Fair WR, Heston MDW, | Israeli RS; |
| DR | WPI: 94-167129/20. | |
| DR | N-PSDB: 065520. | |
| PT | Prostate-specific membrane antigen and DNA encoding it - is | |
| PT | useful for detecting haematogenous micro-metastatic tumour cells | |
| PT | and for identifying ligands which bind to PSM Ag | |

PS Claim 24: Page 103-106: 196pp: English
CC R55097 is a prostatic specific membrane antigen (PSM, encoded by
CC 9055320). The inventors have identified 3 peptides of high
CC hydrophilicity to which antibodies may be directed against. In the
CC specification the specific membrane spanning domain is stated to be
CC from amino acid residue 19 to residue 44. The PSM coding sequence
CC is useful for suppressing or modulating the metastatic ability of
CC prostate tumour cells to grow, or for eliminating them. The protein
CC is useful to identify or purify ligands of the Ag. It is also an
CC attractive target for Ab-directed imaging and targeting of prostatic
CC tumour deposits.
CC Sequence 750 AA:

| | | | | |
|---------------------------|--------|--------------------|-------------|-------------|
| Query Match | 17.2%; | Score 725; | DB 1; | Length 750; |
| Best Local Similarity | 27.3%; | Pred. No. 1.6e-60; | | |
| Matches 213; Conservative | 142; | Mismatches 323; | Indels 102; | Gaps 22 |

| | | | |
|----|-----|---------------------------------------------------------------|-----|
| QY | 65 | RPRPNLIPMAAARBRAPYLVLTAALIFPGALFGLVAFRGSOQAQGSUVLSEBPNV | 124 |
| Db | 17 | RPR-----WLCAG-----ALVLAGFELLGL-----FGWITKSSNEATNI | 52 |
| QY | 125 | EPDLDIFHOGRLYMSDLOAMFLOFEGEGRLEDITROTSLREHVASGAMALITODIRAAIS | 184 |
| Db | 53 | TPK---HNMKAFIDELCK-----ENIKKFLVNFQIPIHAGTEQNPQLAQIOSQW | 101 |
| QY | 185 | RQKLDHWYTDTHYVGLOPPDPANHTLHWVDEACKVGEOLPLEDP-----DYQCPY | 235 |
| Db | 102 | EFGLDSVELAHYDVLSTPNKTHPNYISIIINEDENEIFNTSLEPPPGENVSNDIYPPF | 161 |
| QY | 236 | SAT---GNVTELYEYAHNGREPDIQDI-RANGVPVCRLLLYRNGVISTFOKNTMADFG | 291 |
| Db | 162 | SAFSPQMGPESDLYVYNARTEDFFKLERDKIKCSKATVARTAGKVRGNKKYKMAQLAG | 221 |
| QY | 292 | AQGVLLYPEPADFSODPEKPSLSSQOAVYGHVH-----GTGDPYTPGFPSENOTFP | 344 |
| Db | 222 | AKGYLXSDPADYFAPAGKSVPOGMNLPGGVGQNGNLNLNGADPLTPGIPANEVAYRR | 281 |
| QY | 345 | PVASS-GLPSTPAPDIPISADIASRLRLKRGVAP-QEWQSSLGSPYHLGPG-----P | 395 |
| Db | 282 | GIAENVGLPSPVHPPIGYDDAKOLLEKKMGCSAPPDSSMRGS-LKVPVNVPGFGFNSTQ | 340 |
| QY | 396 | RLRLVNVNHRSTPFINNFCGTECRSEPDHVVYGAORDMGPAAKSAVGTALILETVR | 455 |
| Db | 341 | KVKHHIHSTNVTYIINYIGTLRAVAPEDRYVILGGHDSWFEGCIDPQSGAAVHEIVR | 400 |
| QY | 456 | TESSWVNSGEPRRSRLFTISWDGDFGSGVSTEMLEGYLVHLKAVVYSLDMAVYDGD | 515 |
| Db | 401 | SFGTLRKKGWPRRTIIFASMDAAEFGLGSTTEMAENSRLQGRGAYIINADSSIGNY | 460 |
| QY | 516 | KFHAKTSPLTSLTIESYLKQVDSFNH--SGQTLXEQVYVFNPSMDAEVIRPLMDSSAYS | 573 |
| Db | 461 | TLRWDCPTPLMYSLVHNHLKELKSPDEGEKSLYESWTKKSPSEFGMPRISIKLGSND | 520 |
| QY | 574 | FTAF---VGVPAVESEMEDQ-----AYPLTHKEDTIEYHLNVLOGRLPAAQAQNAQL | 625 |
| Db | 521 | FEVFFORGLAISGARAYTKMWEIKNESGYPLIYSHVETLEYELVEKFDYPMF-KYHLTAQV | 579 |
| QY | 626 | AGOLLIRLSHRLPLDIFGRGADVLRHIGNLNEFS---GDLKARGILTOWYVSARGDY | 681 |
| Db | 580 | RGWVWFELANSIVLPFCRODRAVYLKRYADKIYISAKHNOEMKTYISVSDSLFSAYKNF | 639 |
| QY | 682 | IRAAEKLQEOIYSEEDERLTRYVNRIMKVEFYFLSQVYSPADSPF-RHIFMGRCDH | 740 |
| Db | 640 | TEIASKESERLODQDKSNPIVLRMMNQIMLFEAFIDPLGLP-DRPFYRHVIVAYPSHN | 698 |
| QY | 741 | LGALLDHRILRLRSSSGCPGATISSTGQESR-----FRQDALLTITLOGAANAIS | 791 |
| Db | 699 | KYA-----GESFPGIYDGLDFIESKVDSKAMGEVKKOIRUYAAATVAAAATTS | 747 |

| ID | standard; Protein; 750 AA |
|--------|---------------------------|
| W02234 | |
| W02234 | |

| | | |
|----------------------------------------------------------------------|------------------------------|--|
| 04-NOV-1996 | (first entry) | |
| Prostate-specific membrane antigen. | | |
| Prostate-specific membrane antigen; PSM, promoter; prostate cancer; | | |
| metastasis; therapy; diagnosis. | | |
| Homo sapiens. | | |
| Key | Location/Qualifiers | |
| domain | 20..43 | |
| modified_site | /label= Transmembrane_domain | |
| modified_site | 106 | |
| modified_site | /label= N-glycosylation_site | |
| modified_site | 121 | |
| modified_site | /label= N-glycosylation_site | |
| modified_site | 140 | |
| modified_site | /label= N-glycosylation_site | |
| modified_site | 153 | |
| modified_site | /label= N-glycosylation_site | |
| modified_site | 336 | |
| modified_site | /label= N-glycosylation_site | |
| modified_site | 476 | |
| modified_site | /label= N-glycosylation_site | |
| modified_site | 638 | |
| modified_site | /label= N-glycosylation_site | |
| W09626272-A1. | | |
| 29-AUG-1996. | | |
| 23-FEB-1996; U02424. | | |
| 24-FEB-1995; US-394152. | | |
| 02-JUN-1995; US-470735. | | |
| 02-JUN-1995; US-466381. | | |
| (SLOAN) SLOAN KETERING. INST CANCER RES. | | |
| Fair WR, Heston MDW, Israeli RS. | | |
| Wp1; 96-402365/40. | | |
| DNA encoding alternatively spliced prostate-specific membrane | | |
| N-PSDs; T36785. | | |
| antigen - useful to develop prods. for detecting haematogenous | | |
| micrometastatic tumour cells, or prostate cancer progression | | |
| Example 1; Fig 47A-D; 284pp; English. | | |
| Prostate-specific membrane (PSM) antigen (W02234) is a type II | | |
| integral membrane glycoprotein that is highly expressed in prostatic | | |
| tumours and metastases and is almost entirely prostate-specific. | | |
| It provides an attractive cell surface epitope for antibody-directed | | |
| diagnostic imaging and cytotoxic targeting modalities. Its amino | | |
| acid sequence was deduced from a cDNA clone (T36785) obt'd. from | | |
| human lymph node carcinoma of prostate cells. An alternatively | | |
| spliced variant of PSM (PSM') lacks the first 57 amino acids of PSM | | |
| and is probably cytosolic. Demn. of PSM/PSM' tumour index can be | | |
| used to assess prostate cancer progression; PSM is the dominant | | |
| form in primary prostatic tumours. | | |
| Sequence 750 AA; | | |

Query Match 17.2%; Score 725; DB 1; Length 750;
Best Local Similarity 27.3%; Pred. No. 1.6e-60;
Matches 213; Conservative 142; Mismatches 323; Indels 102; Gaps 22

| | | | |
|----|-----|---------------------------------------------------------------|-----|
| Oy | 65 | RPRQNLIPMAAARBRAFYLVLFALLIFGAFILGVAVRSGQACGDSVLVGEVNY | 124 |
| | 17 | RPR-----WLCGA-----ALVAGFELFGLF-----FGWFIKSSNEATNI | 52 |
| Db | 125 | EPDDLDFHOGRLYMSDLQAMFLOFEGEGLDETJROTJLREVRVAGSAGMAALTODIRALS | 184 |
| Oy | 53 | TPK---HMKRAFDELKRA-----ENIKKFLVNFIQJPHLAGTEQNFLAKIOSQWK | 101 |
| Db | 185 | RKLDHWITDTHIYGLQFPPDAHPNTHLHWDEACKVGEOLPLEDP-----DYVCY | 233 |
| Oy | 102 | EFGDLSVELAHYDVLSTSPKNTHPNYISINEDSNEIENFTSLEPPRPGYENNSDIYVPF | 161 |
| Db | 236 | SAI---GNVTGELVYAHARGREDIODI--RARGVPVGRLLVLRGVISFAQKVTNAODFG | 291 |
| Oy | 162 | SAFSQGMPEDDLYVYNARJEDFETKLERDMKIKCSCKIYIARIGKVFRRGNKAKYNNOLAG | 221 |
| Db | 292 | AGVLYLPEPADFSODPPKPSLSQQAUVYGHVH-----GTGDPYTPGPFSENFOTFP | 344 |

KW fusion protein; prostate specific membrane antigen; immunofusin;
 KM extracellular domain.
 OS Homo sapiens.
 PN US572604-A.
 PD 10-MAR-1998.
 PF 14-SEP-1995; 528122.
 PR 14-SEP-1995; US-528122.
 PR 14-SEP-1994; US-305700.
 PA (FUT-) FUJI IMMUNOPHARMACEUTICALS CORP.
 PI Gillies SD, Lo K, Sudo Y;
 DR WPI; 98-192768/17.
 PT Recombinant DNA for expression of target protein, e.g. HIV gp120 -
 PT comprises sequences coding for signal peptide, immuno-globulin Fc
 PT region and gp120
 PS Example 14: Columns 27-32; 18pp; English.
 CC This is the extracellular domain of the prostate specific membrane
 CC antigen (PSMA) which is a type II membrane protein. The extracellular
 CC domain is modified to construct a new recombinant fusion protein, a
 CC PSMA immunofusin. The new recombinant DNA construct is used for
 CC expression and secretion of a target protein, whose sequence is free of
 CC immunoglobulin CH1 domain. The construct comprises a polynucleotide
 CC encoding from its 5' to 3' direction a secretion signal sequence,
 CC comprising a sequence encoding an immunoglobulin Fc region, and a
 CC sequence encoding the target protein. A replicable expression vector
 CC comprising the polynucleotide sequence can be used to transfect a host
 CC cell. The products can be used to produce a recombinant fusion protein
 CC (immunofusin) comprising the Fc region and target protein. The DNA can
 CC be expressed at high levels in a host cell, and the fusion protein is
 CC efficiently produced and secreted.
 SQ Sequence 707 AA;

Query Match 16.7%; Score 702.5; DB 1; Length 707;
 Best Local Similarity 28.4%; Pred. No. 2e-58;
 Matches 193; Conservative 128; Mismatches 291; Indels 67; Gaps 17;
 QY 166 VASASAGMAALTODIRALSNQKIDHWTDHYVGLQFPDDAHPTLWVDEAGKVGSDLP 225
 Db 40 LAGDONFOLAKOIQOSWKKEGDLSDVELAHVDLSTPNKTHPYISITINDEGNEIFNTS 99
 QY 226 LEDP-----DYVCPSNAI--GNVTGELYVAHGREDDLOD-L-RARGVDVGRLL 272
 Db 100 LFEPPREYGENVSDIVPEFSAFSQGMPEGDLYVNAKTEDEFKLERDKINSGKIV 159
 QY 273 VRVGVISFAQKVTINADQFGAGVLIPEPADFSQDPKPSLSQAAVYGHVLT----- 326
 Db 160 ARRGKVRGNKKVKAQLAGAKGVILYSDPADYFAVPKVPDGNVLPGGQVGRNITLNLN 219
 QY 326 GTDDPTYPGPFSPFNOTQFPYPASS-GLPSTPAOPISADIASRLRLKLGYPAP-QEOMGS 383
 Db 220 GAGDPLTPGYPANETAYARRGIAEAVGLPSIPVHPHIGYDAOKLEKKGSAFPDSSWRGS 279
 QY 384 ILGSPYILGGG-----PRLRLVNNHRTSTPINNIFGCIEGRSEPDHYVIAQORDAW 436
 Db 280 -LKVPIYVGGEGFTGNSTQKAKHISTNEVTRITYNVIGTLRGAVEPDRYITILGCHDSW 338
 QY 437 GPGAASAVGCTAILLELVRTSSMVSNGFRRRSILFIISMDGGDGSVSTEWLEGYLSV 496
 Db 339 VFGIDPQSGAAVYHEIVRSGLTKKEGWRPRRTILFASMDAEEFGILSTEMAENSR 398
 QY 497 LHLKAVVYVSLDANVLDDDKFHAKTSPILSLIESVYLKQVDSPNH--SGQTLISQVYFTN 554
 Db 399 LOEGGVAYINADSSIEENYLLRVDCPLMWSLVNHLTKELKSPDEGEKSLYSMTKRS 458
 QY 555 PSMAEYIRPLPMDSASVSTFAF--VGVPVFEFSFMDDD-----AYPFLTKEDYEN 606
 Db 459 PSPEFSGMPRTSKLGSNDPEVFQFQGIASGRARYTKMNETKNFSGYPLHSHYETTEL 518
 QY 607 LHKVLQGRLPAAVAQAQLQGLLIRLSHDLRLPLDFGRYGVYVLRHIGNLNEFS---G 662
 Db 519 VEKTYDPMF-KYHLTAQVRCGMVFEIANSIVLPFCRDYAVVLRKYADKLYTSKMKHPQ 577
 QY 663 DKARGLTLOWNVYARGDYIAAEKRLQELTISSEDERELTRMTNVRIMRVEFFLSQYV 722

Db 578 EMKTVSVSDSLFSAYKNFTEISAKFSERQDDPKSNPIVLRMANNQMLMELEAFIDPLG 637
 QY 723 SPADSEF-RHIFMGDGHFLGALLDHLRLRSNSSGTSGATSTGFOESR-----F 772
 Db 638 LP-DREFYRHVIVYAPSSHNKYA-----GESFPGIYDALFDIESKVDPSKAMGEV 685
 QY 773 RQDALITWTLOGAANALS 791
 Db 686 KROIYVAAFVTOAAETLS 704
 RESULT 7
 ID W33604 standard; Protein: 472 AA.
 AC W33604;
 DT 21-MAY-1998 (first entry)
 DE Human secreted protein AM282 full-length sequence.
 OS Secreted protein; AM282; cytokine; human.
 FH Homo sapiens.
 FT Key Location/Qualifiers
 FT Peptide 1..24
 FT Protein /label= Sig_peptide
 FT 25..472 /label= Mat_protein
 PN W0979030-A2.
 PD 23-OCT-1997.
 PF 16-APR-1997; 006475.
 PR 13-JAN-1997; US-783520.
 PR 18-APR-1996; US-634325.
 PA (GENEY) GENETICS INST INC.
 PI Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racie LA,
 PI Spaulding V;
 DR WPI; 97-526400/48.
 DR N-PSDB: V02296.
 PT New isolated secretory proteins AM340, AM282 and AK583 - possibly
 PT have cytokine, cell proliferation/differentiation regulating,
 PT immunomodulating activities, etc.
 PS Claim 16: Page 45-47; 59pp; English.
 CC This human secreted protein, designated AM282, is encoded by a
 CC full-length cDNA clone (see V02296), deposited in ATCC 98026, that
 CC was identified from a database search using an isolated partial
 CC AM282 clone (see T97398). AM282 protein can be used in a claimed
 CC method for preventing, treating or ameliorating a medical
 CC condition. It may exhibit cytokine, cell proliferation (either
 CC inducing or inhibiting) or cell differentiation (either inducing or
 CC inhibiting) activity or may induce production of other cytokines in
 CC certain cell populations. It may also exhibit e.g. immune
 CC stimulating or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic or chemokinetic activity, haemostatic or thrombolytic
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC tumour inhibition activity, or other activities. No evidence of
 CC any of these activities is given in the specification.
 SQ Sequence 472 AA;

Query Match 3.3%; Score 140; DB 1; Length 472;
 Best Local Similarity 17.8%; Pred. No. 9.3e-05;
 Matches 106; Conservative 92; Mismatches 216; Indels 182; Gaps 25;
 QY 93 FTGAFLIGVAFRSGCO-----ACGDSVLVSEVDVNEPDLDFHGRGLY 136
 Db 9 FCGVHLLSLCSGAKICXNGKISKRTFEIKEIKELISCCGVARAIIILVAYGR---AQNRSY 64
 QY 137 WSDQANFLOFLGREGREDIIRQTSLEBRVAGSGMAALNODIRAAISROKLDHWMTDT 196
 Db 65 --ERLALVDTVG--*-----PRLSGSKNEKAIOIMQONQDQGLEVY-----H 104
 QY 197 YVGLOFPDPAHPNPLIHWVDAKGVEQLPLEDDVNYCPY-----SAIG-----NVTGELYV 247
 Db 105 LEPIRIP-----HW--ERGE--ESAVMLEPRIRHIALIGLSSIGTPEEGITAEV 152

```

OY 248 AHYGRPELDQDARGVDPVGRLLVAVGVISFAQV-----INADFGAGVLIYPE 300
DB 153 V-----TSFDELQRRASEGRKIVVYNODYINYSRTVOYRTGGAVERAAGALASLI--RS 206
OY 301 PADFSODPPKPSLSSQQAAYGVHVLGTGDPYTPGPFSENOTQFPVASSGLPSIPAQIS 360
DB 207 VASFSTIYSPHTIOGYO-----DGVKIPITACT 235
OY 361 ADIASRLKLKLPVAPDQEMOGSLGSPYHLGPGPRLRLVNNHRTSPINNIFSCIEGR 420
DB 236 VEDAEKMSHM-----ASHGIRIVLIQKKMGAKTPTDTSFNTV--AEITGS 278
OY 421 SEPDHYVYVGAQRDAW--GPGAASAVGTAILLELVRTSSAVNSGFRPRRSLLTISMDG 478
DB 279 KPEOVYVLSGHLSDWVGQGGAMDDGGAFISWEALSLIKDI--GLRKRRLRLVLMWA 335
OY 479 GDFGVSGSTEM-----LEGYLVHLKAVYVSLDNAVLGDDKFRHAKTSPLTSLIES 531
DB 336 GGGGVGAFQYQVLHKVNSNSLYMESDAGTFLPTGLQFTSEKRA-----IMGE 387
OY 532 VLKQVDSPNHSGQTLVEOVVFTNPSMDAEVIRLPM-----DSSAVSFAVGVPAVE 584
DB 388 VM-----SLDPLNITQVLSHGEGDINFWIQAGVPGA- 421
OY 585 FEFMEDDAQPELHTKEDTYENLHKVLOGRLPAVQAQVLAQGLLIRLSHRLLP 640
DB 421 -SLDDLYIKYFFEH---HSHGDTMTVMDPKMNVAAMVAVSYVADM--EEMLP 470

```

```

RESULT 8
ID R99416 standard; Protein: 393 AA.
AC R99416.
DE 04-DEC-1996 (first entry)
DE Amino peptidase precursor of Aeromonas caviae, useful in food mfr.
KW Amino peptidase; hydrophobic amino acid removal; hydrolysis;
KW food manufacture; improved taste.
OS Aeromonas caviae T-64.
FH Key Location/Qualifiers
FT 121.393
FT 121.393 /label=mat_peptide
FT J08173168-A.
PN 09-JUL-1996.
PD 09-JUL-1996.
PF 26-DEC-1994; 336663.
PR 26-DEC-1994; JP-336663.
PA (NORQ) NORINSUISANSHO SHOKUHN SOGO.
PA (ZENK-) ZENKOKU RAKUNOGYO KYODO KUMIAI RENGOKAI.
DR WPI: 96-365587/37.
DR N-PSDB: T41568.
PT DNA encoding aeromonas microorganism derived aminopeptidase - which
PT eliminates bitter tastes by decomposing peptide(s) rich in
PT hydrophobic amino acids produced by protein hydrolysis, useful in
PS food mfg.
PS Claim 1: Page 5-7; 12pp; Japanese.
CC R99416 is an amino peptidase precursor from Aeromonas caviae. The
CC peptidase is useful in the hydrolysis of peptides high in hydrophobic
CC amino acids, this is useful in food manufacture as such peptides have
CC a bitter taste. The amino peptidase removes amino acids sequentially
CC from the N-terminus of a peptide at an optimum reaction temperature
CC of 50 deg. C. and a pH of between 8 and 10.
SQ Sequence 393 AA:

```

```

Query Match 3.1%; Score 132; DB 1; Length 393;
Best Local Similarity 20.9%; Pred. No. 0.0004;
Matches 96; Conservative 51; Mismatches 150; Indels 162; Gaps 18;

```

```

OY 256 IQDLRARGVDPYGRLLVAVGVISFAQVYTNADFGAGVLIYPEPADSODPPKPSLSS 315
DB 12 LSPLAQAQAEPPWITVGADSGVELKQVKARLAPLFSASAPQLQAVESELGTLISHLMH 71
OY 316 QQAVGVHVLGTGDPYTPGPFSENOTQFPVASSGLPSIPAQISDIAS----- 366

```

```

DB 72 E-----GHRGCG-----GYVHS-----TLADALQSM-AQIPSONLFSAPPLTQASV 113
OY 366 -RLLRKLKGPVAPDQEMOGSLGSPYHLGPGPRLRLVNNHRTST----- 409
DB 114 NRLT-----PYLDQGNIVGTISQLASMR-----NRYTTTGTQOSADWAGQMSL 159
OY 409 -----PINNIFGICGRSEPDHYVYVGAQRDA-----WGGA 441
DB 160 SATLPWASVSKYKSGYPPQGSVLTLLKGSRYPDDEVVYVGLDSTAGSAPNSRTLAPAD 219
OY 442 KSAVGTAILLELVRTSSAVNSGFRPRRSLLFIEMDGGDFGSYSTEWLEGYLVHLKA 501
DB 220 DDASIALTLTEVLAV---IAQGRQPERTLOFTGAAEVGLRSKDLA----- 266
OY 502 VVYVSLDNAVLGDDKFRHAKTSPLTSLIESVYLKQVDSPNHSGQTLVEOVVFTNPSMDAEV 561
DB 266 -----TRYKANTRVLAAL-----OLDMTNYGSA--EDIVMTDYDQGL 304
OY 562 IRPLMDSASFTAFVGVPAVESEFME-----DQOAVP----- 597
DB 305 TGYLAQLLDAY-----LPQIRGYDSCGYGCSDHASWHQGYPAAMPFESRENDYNPK 357
OY 597 LHTKEDTYENLHKVLOGRLPAVQAQV--AQLAGOLLIRL 633
DB 358 IHTADT-----LQNSDPSAHLKRFQALATSFATFM 389

```

```

RESULT 9
ID W85456 standard; Protein: 472 AA.
AC W85456.
DE 25-FEB-1999 (first entry)
DE Secreted protein encoded by clone bu45.2.
KW Secreted protein; nutritional activity; immune stimulating; vaccine;
KW suppressing activity; haematopoiesis regulating activity;
KW tissue growth activity; activin; inhibin activity; chemotaxis;
KW chemokine activity; haemostasis; thrombolytic activity; receptor;
KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
KW tumour inhibition; gene therapy.
OS Homo sapiens.
PN WO9842739-A2.
PD 01-OCT-1998.
PR 20-MAR-1998; U05653.
PR 19-MAR-1998; US-044466.
PR 21-MAR-1997; US-822167.
PA (GENY) GENETICS INST INC.
PA Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racine LA, Spaulding V, Treacy M;
DR WPI: 98-609890/51.
DR N-PSDB: W82779.
PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, placenta or adult
PT pineal gland cDNA libraries.
PS Claim 14: Page 70-72; 11pp; English.
CC The present sequence represents a secreted protein. The polynucleotide
CC and secreted protein are predicted to have biological activities which
CC would make them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is given.
CC suggested activities include nutritional activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, haemostatic and thrombolytic activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity (no data is
CC given in the specification to support these activities). The
CC polynucleotide is also stated to be useful for gene therapy.
SQ Sequence 472 AA:

```

```

Query Match 3.1%; Score 132; DB 1; Length 472;
Best Local Similarity 17.6%; Pred. No. 0.00054;
Matches 105; Conservative 92; Mismatches 217; Indels 182; Gaps 25;

```

```

OY 93 FTGAFLGTVAFRSCQ-----ACGDSVLVSEDEVNPEBDLDFHGRRLY 136
    |||
DB 9 FCGVHLILSCSGKAICKNISKRTFEEIKETASCGDVAKAINLAVYK---AQNRSI 64
OY 137 WSDLOAMFLOFLGEGRLIEDTIROTSREVRVAGSAGMAALTODIRALSRKDLHWTDTH 196
    |||
DB 65 --ERLALLVDVYG-----PRLSGSKNLEKAIQIMYQNLQDDGLEK----H 104
OY 197 YVGLQFPDPAHNTLHWDEACKVGEQLPLEDDVYCPY-----SAIG-----NVTGELY 247
    |||
DB 105 LEPVRIP-----HM--ERGE--ESAVMLEPRHKIKAILGLGSSIGTPPEITAEVLY 152
OY 248 AHYGRREDLODRLRARGVDVGRLLVVRGVISFAQVY-----TNAQDFGAGVLIYPE 300
    |||
DB 153 V-----TSTDELDORRASEARKIIVYNOPYINSTRVOYRTQCAVFAKAVGALASLI-RS 206
OY 301 PADFSODPKPKPSLSQQAAYGVHVLGTGDPYTPGFPSFNOTOPPPVASSGLPSIPAPIS 360
    |||
DB 207 VASFSTSPHTGIGQEQY-----DOVKPIPTACIT 235
OY 361 ADIASHLKRLKGPVAPQEMQSGSLGSPYHLGPRRLRVNNHRTSTPINNIEGCIER 420
    |||
DB 236 VDAEAMSM-----ASHGIKIVIQLKMGAKTYPTDSEFNTV--AEITGS 278
OY 421 SEPDIYVVGARDAW--GPGAASAVGTAILLELVRTSSMWSNGFRRRSLIFSMG 478
    |||
DB 279 KYPEOVYLVGHLSDMDVCGADGGAFISWALSLIKDL--GLRKRFLRLVLMWA 335
OY 479 GDFGVSSTEW-----LEGYLSVLKRAVYVSLDNVAGDDKPHATSPLLTSLIS 531
    |||
DB 336 EEOGGVGARQYQYQHLKVNINSYSLVWESDAGTFLPTGLQFTGSEKRA-----IMEE 387
OY 532 VLKQVDSFNHSQOTLYEQVFTNPESWDAVIRLPM-----DSSAYSFTEAVGVPAVE 584
    |||
DB 388 VM-----SLQPLNITQVLSHGEGTDINFWIQAGVGA- 421
OY 585 FEFMEDDAVPLHTKEDTYENHKVLOGRLPRAVQAQVAGOLLIRISHRLP 640
    |||
DB 421 -SLDDLYKYFFFH--HSHGDTMTVMDPKMNVAAMVAVVSYYADM--EEMLP 470

RESULT 10
R27481
AC R27481 standard; Protein: 806 AA.
DE 03-MAR-1993 (first entry)
DE RP-III residual protease.
KW mutation; vpr; cleaning; laundry; detergent; washing powder; apr;
KW npr; epr; bpr; rp-I; mpr; rp-II; spooA gene;
KW reduced sporulation dependent protease.
OS Bacillus subtilis.
PN WO9216642-A.
PD 01-OCT-1992.
PE 26-FEB-1992; U01598.
PR 19-MAR-1991; US-671376.
PA (OMNI-) OMNIGENE INC.
PI Peto J, Ruto GA, Sloma A;
DR N-PSDB: Q29134.
PT Bacillus residual protease III (RP-III) gene - which can be
PT mutated for prodn. of heterologous polypeptide(s) or expressed to
PT produce protease
PS Claim 25; Fig 4; 28pp; English.
CC This sequence represents residual protease III (rp-III) from
CC Bacillus subtilis, and was decoded from the appropriate DNA as detailed
CC in Q29134. The protease is secreted by B. subtilis in large amounts, and
CC can cause proteolytic degradation of any protein produced by the
CC bacteria, ie. a recombinant product. By introducing mutations in the
CC protease gene heterologous polypeptides can be expressed which do not
CC cause lowered levels of proteolytic damage. Also RP-III protease can be
CC expressed for use in improving the cleaning activity of laundry powders
CC or for use in industrial processes.
SQ Sequence 806 AA:

```

```

Query Match 3.1%; Score 130; DB 1; Length 806;
Best Local Similarity 21.8%; Pred. No. 0.0021;
Matches 101; Conservative 52; Mismatches 163; Indels 148; Gaps 23;

OY 81 AAPYLVTLLFLFTGAFLLGYAFRSCQACGDSVLVSEDEVN-EPDL--DFHGRRLY 137
    |||
DB 164 SAPYIGA-----NDAMDLGYT-----GKGIKVAIIDTGVYHNPDLKKNFGYKGY- 210
OY 138 SLOAMFLOFL-----GEGRLIEDTIROTSREVRVAGSAGMAALTODIRALSRKDLHWTDTHYV- 171
    |||
DB 210 -----DFVNDIDPKRTPTGDPREGATBDGTHAGVAAANGTIKGVAPDAILAYRY 261
OY 171 ---GMAALTODIRALSRKDLHWTDTHYV--LOFPDPAHNTLHWDEAKV----- 221
    |||
DB 262 LGPGSGTTEVYIAGVERAVQDADVMNLSTNSLNNPDMATSTALDMMSEGVAVTSN 321
OY 221 -----GEOLPLEDDV-YCPIS-----AIGNVT 242
DB 322 GNSGPNMTVGSPTGSRSAISVGAQQLPNEYAVTFGSYSAKVMGYNKEDDYKALNKE 381
OY 243 GELVYAHYGRPDLODLRARGVDVGRLLVVRGVISFAQVY-----TNAQDFGAGVLIYPE-P 301
    |||
DB 382 VELVEAGTEADFE-----GKDLGKVAVVRKGSIAFYDKADNKKAGATGVYNNIS 436
OY 302 ADFSODPP-----KPSLSQQAAYGVHVLGTGDPYTPGFPSFNOTOPPPVASSGLPSI 354
    |||
DB 437 GELEAVNPMGVSPTIKLSLEDEKLYSALKAGE-----TKTFKLYSKAL---- 483
OY 355 PAQPSIADIASHLKRLKGPVAPQEMQSGSLGSPYHLGPRRLRVNNHRTSTPINNIF 414
    |||
DB 483 -GEQV-ADEFSR-----GPVM-DTW---MIKPDISAGVNIIVSTIPTHDDHPG--Y 527
OY 415 GCIEGRSEPDHYV-----VIGAQRDAWGPAAKSAV-GTAILLE 452
    |||
DB 528 GSKQGTSMASPHIAGAVAVIKAKPKWSVEQIKAIMNTAVTLK 571

RESULT 11
Y04867
AC Y04867 standard; Protein: 500 AA.
DE 06-JUL-1999 (first entry)
DE Mycobacterium species protein sequence 19D.
KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection.
OS Mycobacterium sp.
PN WO909186-A2.
PD 25-FEB-1999.
PE 14-AUG-1998; F01813.
PR 11-SEP-1997; FR-011325.
PR 14-AUG-1997; FR-010404.
PA (INSP) INST PASTEUR.
PI Gicquel B, Llm Em, Pellicic V, Portnoi D, Gognet de la Salmoniere Y,
PI Guigueno A;
DR WPI: 99-181045/15.
DR N-PSDB: X34119.
PT Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in
PT infection-associated protein expression
PS Claim 32; Fig 19D; 309pp; French.
CC Sequences Y04742-Y05000 and Y07201-Y07204 represent secreted proteins
CC from various Mycobacterium species microorganisms. The encoding
CC nucleotide sequences can be used as primers and probes for methods
CC for detecting and identifying mycobacteria, especially belonging to
CC the M. tuberculosis complex. The encoded proteins can be used in
CC vaccines for immunisation against a bacterial or viral infection.
SQ Sequence 500 AA:

```

Query Match 2.9%; Score 120.5; DB 1; Length 500;
 Best Local Similarity 23.3%; Pred. No. 0.0076;

[illegible]

| | | | | |
|-----------------------|--------|-------------------|-------|-------------|
| Query Match | 2.98; | Score 120.5; | DB 1; | Length 527; |
| Best Local Similarity | 23.38; | Pred. No. 0.0083; | | |

Matches 125; Conservatively 55; Mismatches 186; Indels 171; Gaps 30;

```
QY 161 SLREIVASSAGMAALT--ODIRALSL-----ROKIDHWTDTHHYGLQFPDPAPR 208
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 74 ALRNRVSTIDMAHLSTKQDIDIANNDGIRAVGPGQASVDVYVNTLRNSFDVQTPEFS 133
QY 209 NTLHWVDE-----ACKVGEQLPLE-----DPDYICPTSAIGNVTGELYAHYG----RP 253
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 134 ARVFAKEGVVTLTGNTVEARALEVSLTGTPD-----GVNCPVLAAPADSPGCSFP 184
QY 254 EDLDQDLRARGVDPV-GRLLIVRGVIGISPAQKVTNAQDFGACQGLTYLEPPAPFSDDPKPS 312
Db 185 SDYDPL-----PVSQAVALVDRCGCPFAQKEDDAAGRAVALTI----AD----- 226
QY 313 LSSQQAIVYGHVHLTGDPYTPGPFPSFNOTOPPPVASSGLSPISPAQISADIASRLRLKLK 372
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 226 NIDEGAMGTIGIANDT-----YKIPVSVTKSKVGFQ-LNGQS 261
QY 373 GPVAQEMOGSLSGSPYHLPGPRLRLVYNNHRTSTPINNIIGCTIEGRSEPDHYVVGQAQ 432
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 262 GPTTVK-----LTASTQSF--KARNVIAQTKTGSSAN-----VYMGAGH 298
QY 433 RDA--WGRGAAKSANGTALILELVRTF--SSMWNGFRPRRSLEFISMDGDSFGSVSTE 488
Db 299 LDSVEPGEGINDNGSGVAAYLETAQIGNSPHVSNAYK----FATMGAEFEFLISRN 352
QY 489 WLEGYLSVHLTKAV-VVYSLDNAYV-----GDQK--FRAKTSPLT--TSLIE-- 531
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 353 YVES-LDIDLAKGIALYINFDMLASPNRGYTFYQDQSLPLDAAGQPPVPGSAGIERTF 411
QY 531 -SVLKQVDSPNHSQGLTYEQVVFETNPSSMDAEVIRPLPMDSSAYSFTFAVGPAVE-FSFM 588
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 412 VAYLKXK-----AGKTAQD-----TSPDGR-----SDYGGFTLAGIPSGSLFSGA 450
QY 589 E-----DDQATPFLTKREDTYENLHKVLQGLRPA-VAQAVALAGLQGL 629
Db 451 EVKSKAEQAELMGSTADEPFDPNYHOKTDLTDLIDRTALGINGAGVAAVGLYADQL 507

RESULT 13
W89614
ID W89614 standard; Protein; 496 AA.
AC W89614.
DT 17-MAR-1999. (first entry)
DE Aspergillus oryzae dipeptidyl aminopeptidase #2.
KM Prolyl dipeptidyl aminopeptidase; protein hydrolysate; dough;
      flavour enhancer; palatability; mouthfeel; aroma; crust colour;
      baking; animal feed additive; hydrolysis.
OS Aspergillus oryzae.
PN W09851803-A1.
PD 19-NOV-1998.
PF 12-MAY-1998; U09629.
PR 20-OCT-1997; US-062892.
PR 16-MAY-1997; US-857884.
PI (NOVO ) NOVO NORDISK BIOTECH INC.
PI Blinkovsky A, Brown K, Bynun T, Klotz A, Rey MW;
DR WPI; 99-045232/04.
DR N-PSDB: X00070.
PT New dipeptidyl aminopeptidase from Aspergillus oryzae - used to
PT produce protein hydrolysates enriched in particular amino acids,
PT useful as flavour enhancers, e.g. in doughs
PS Claim 1; Fig 1; 77pp; English.
CC The present sequence represents dipeptidyl aminopeptidase (DPAp) from
CC Aspergillus oryzae. DPAp acts synergistically with an aminopeptidase
CC (AP) to hydrolyse polypeptides, producing protein hydrolysate (PH),
CC useful in foods as flavour enhancer, e.g. in baked goods, enriched in:
CC (a) Ala, Arg, Asp, Gly and/or Val, or (b), if the substrate has been
CC deamidated, in Glu (free and/or peptide bound), in which case products
CC are useful as animal feed additives. DPAp can also be used in flavour-
CC improving compositions (optionally containing AP) and in dough pre-mixes,
CC also for deactivating enzymes and for converting precursors to mature
CC proteins. DPAp increases the level of hydrolysates of proteins and thus
CC of flavour development, and a mixture with AP may hydrolyse tripeptides
```

CC also involves deactivating enzymes and/or converting precursors to mature
CC proteins. DPAH increases the level of hydrolysis of proteins and thus
CC of flavour development, and a mixture with AP may hydrolyse tripeptides
CC of flavour development, and a mixture with AP may hydrolyse tripeptides

CC that are resistant to either enzyme used alone. PH have improved
CC solubility, emulsifying and foaming properties, and products containing
CC them have better flavour, palatability, mouthfeel, aroma and crust
CC colour.
SQ Sequence 496 AA;

| | | | | |
|-----------------------|-----------------|-----------------|------------|------------|
| Query Match | 2.8% | Score 117.5 | DB 1 | Length 496 |
| Best Local Similarity | 20.1% | Pred. No. 0.015 | | |
| Matches 108 | Conservative 77 | Mismatches 194 | Indels 157 | Gaps 28 |

[illegible]

CC foods as flavour enhancer, e.g. in baked goods, enriched in: (a) Leu,
CC Gly (especially, Glu, Ser, Asp, Asn, Pro, Cys, Ala and/or Gln, or (b),
CC if the substrate has been deamidated, in Glu (free and/or peptide bound)
CC in which case products are useful as animal feed additives. AP can be
CC used in flavour-improving compositions (optionally containing EP) and in
CC dough pre-mixes. Also AP can be used for deactivating enzymes and for
CC converting precursors to mature proteins. AP increases the level of
CC hydrolysis of proteins and thus of flavour development. PH have improved
CC solubility, emulsifying and foaming properties, and products containing
CC them have better flavour, palatability and aroma.

| | | | | |
|-----------------------|-----------------|-----------------|------------|------------|
| Query Match | 2.8% | Score 117.5 | DB 1 | Length 496 |
| Best Local Similarity | 20.1% | Pred. No. 0.015 | | |
| Matches 108 | Conservative 77 | Mismatches 194 | Indels 157 | Gaps 28 |

```

OY 139 DLQAEFLQFLGSG--RLDETITQOTSLRFRVAGSAG----MALTFODIRA-----LSROKL 188
Db 28 DIQ----LEDDLEGSQQLDEFAAYPERKRNVRVGGKRAHDVTANLYLELKTGYDYDYKQOP 84
OY 189 DHVWTDTHYVGLQFPDPAPHPTLHWMDAGKAYEOLPLEDDPVYCPYSAGVNGELY- 248
Db 85 VHLMSN-----ADQTLKGDGDE---EIEAKTMTYSVSVYTLADVAV 122
OY 248 -----AHYGRPELODLRAGVDPVGRLLVRYGVTSFAQKYNADQFGAQVLIPE 300
Db 123 KNLGCEADY-PSVYE-----GKVALIKRGCEPFGDKSVLAAKAKAASTIVYNN 170
OY 301 PADFSQDPKPSLSSQQAUVYGHVLGTGDDPYTPGFPFSNQTOFPVASSGLPSIPAOPIS 360
Db 171 VA-----GSMAGTLCMAQSDKG-PYS-----AIVG-----IS 196
OY 361 ADIARLLRKLLK-GEVYAOEQMGSLSGPSYHLGPGPRLRLVYNNHRTSPINNIGCTEG 419
Db 197 LEDGOKLLKLEAGSVSDLM-----VDSKOENTTYNNVVAQTKG 236
OY 420 RSEPHYVYVIGAQBPAMGPAGAKSAVGTALILELVRTFESSMVSNGFPRPSRLSLFTSMDCG 479
Db 227 -GDPNNVVALGGHTDSVEAGPGINDGSGIINLV---IAKALTYYSKNAVRFLEWTA 292
OY 480 DFGSVGSTEWELEGYSVTLKAV-VYVSLD-----NAVU-----GDDKFRAKTSPL---L 525
Db 293 EFGLLGSNNYV-SHLNATELNKIRLYLNFDMIASPNYALMIYDGGGSAFNOSGPGASAOI 351
OY 526 TSLISVYLKQVDSNP-----HSGOTLYEQOVFTNPMSDAEVIRPLPMDSSAYSFTEAVGP 581
Db 352 EKLFEYDYSIDLPHIPIQFQDGRSYE-----AFILGIP---SGGLFTGAEGLM 398
OY 582 AVESFFMEDDQA-----YPLTKREDTYENL-HKVLOGRLPAVAQAVALQAGOL 629
Db 399 SEENASRWGCGAGVAYDANY---HAAGDNMTVLNHEAFLINSKATAFAVATYANDL 451

RESULT 15
ID W89586 standard; Protein; 496 AA.
W89586
AC 17-MAR-1999 (first entry)
DE Aspergillus oryzae aminopeptidase II.
KW Aminopeptidase; protein hydrolyase; glycine releasing; protease;
KW proteolaceous material; flavour; food; baking; animal feed additive;
KW palatability; hydrolysis; solubility; emulsifying; foaming; aroma;
KW mouthfeel; crust colour.
OS Aspergillus oryzae.
PN WO9851163-A2.
PD 15-NOV-1998.
PF 15-NOV-1998. 009998.
PR 16-DEC-1997; US-069719.
PR 16-MAY-1997; US-857886.
PR 20-OCT-1997; US-062893.
PR 16-DEC-1997; DK-001465.
PA (ASAHI ) ASAHI CHEM IND CO LTD.

```


PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 PI Blinskovsky A, Brown K, Byun T, Fujii M, Golightly E,
 PI Kofod LV, Marumotac, Mathiansen TE;
 PI Wpi: 99-045177/04.
 DR N-PSDB: V82521.
 PT Production of protein hydrolysate - using protease and enzyme that
 PT releases glycine, useful as flavour improvers in foods and animal
 PT feed additives
 PS Claim 4; Fig 1; 84pp; English.
 CC A method has been developed for the production of protein hydrolysates
 CC (PH) comprising reacting a protein with: (i) at least one polypeptide
 CC with gly-releasing activity; and (ii) at least one other protease so
 CC that the amount of Gly produced is greater than when the protease is
 CC used alone. PH are used to improve flavour of foods (e.g. baked goods)
 CC and as animal feed additives. PH are preferably also enriched in Glu
 CC (free and/or peptide bound). So have improved flavour and palatability
 CC Addition of a polypeptide with Gly-releasing activity increases the
 CC degree of hydrolysis (or reduces the amount of enzyme needed) and
 CC hydrolysates have better solubility, and emulsifying and foaming
 CC properties. Baked goods containing them have improved aroma, mouthfeel
 CC and crust colour. The present sequence represents *Aspergillus oryzae*
 CC aminopolypeptidase II, which is used in the method of the invention.
 Q0 Sequence 496 AA;

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2000, 22:32:49 ; Search time 13.63 Seconds
(without alignments)
2772.019 Million cell updates/sec

Title: US-09-358-755-1
Perfect score: 4203
Sequence: 1 MERLKGFLFQRAQQLSPRSSO.....TLQGANALSGDVWNIDNPF 801

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database: PIR-62.*

Word size: 0

Number of hits that pass the threshold: 142080

1: pirl:*
2: pirl:*
3: pirl:*
4: pirl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----------|---------------------|
| 1 | 1504.5 | 35.8 | 757 | 2 A48592 | transferrin recept |
| 2 | 1496.5 | 35.6 | 760 | 1 JXHU | transferrin recept |
| 3 | 1450.5 | 34.5 | 763 | 1 S29548 | transferrin recept |
| 4 | 1428.5 | 34.0 | 776 | 1 JH0570 | transferrin recept |
| 5 | 1424.5 | 33.9 | 622 | 2 A34549 | transferrin recept |
| 6 | 725 | 17.2 | 750 | 2 A56881 | prostate-specific |
| 7 | 438.5 | 10.4 | 811 | 2 S57149 | probable membrane |
| 8 | 297.5 | 7.1 | 783 | 2 S65188 | secretory protein |
| 9 | 243 | 5.8 | 809 | 2 S67153 | probable membrane |
| 10 | 208 | 4.9 | 63 | 2 S68317 | transferrin recept |
| 11 | 182 | 4.3 | 455 | 2 S39663 | aminopeptidase hom |
| 12 | 143 | 3.4 | 504 | 2 S24314 | bacterial leucyl a |
| 13 | 136 | 3.2 | 561 | 2 E75191 | hypothetical prote |
| 14 | 131.5 | 3.1 | 565 | 2 E71223 | hypothetical prote |
| 15 | 130 | 3.1 | 806 | 2 A41341 | microbial serine p |
| 16 | 123.5 | 2.9 | 495 | 1 P2WLB2 | L2 protein - bovin |
| 17 | 120.5 | 2.9 | 500 | 2 H70629 | probable AMINOPEPT |
| 18 | 120 | 2.9 | 1448 | 2 T08526 | DNA primase tracr2 |
| 19 | 120 | 2.9 | 1448 | 2 S37669 | tracr2-protein - E |
| 20 | 117.5 | 2.8 | 374 | 2 S69699 | hypothetical prote |
| 21 | 116.5 | 2.8 | 1301 | 2 S18118 | alpha-amylase - Al |
| 22 | 116 | 2.8 | 1017 | 2 B70985 | probable polyketid |
| 23 | 115.5 | 2.7 | 640 | 2 S62747 | homeotic protein A |
| 24 | 113.5 | 2.7 | 634 | 2 T00388 | hypothetical prote |
| 25 | 112.5 | 2.7 | 701 | 2 F48613 | gag polyprotein - |
| 26 | 112.5 | 2.7 | 701 | 2 D48613 | gag polyprotein - |
| 27 | 112.5 | 2.7 | 1603 | 2 A48613 | gag/pol polyprotei |
| 28 | 112.5 | 2.7 | 1805 | 2 A44736 | nestin - rat |
| 29 | 112 | 2.7 | 894 | 1 A41527 | protein-tyrosine k |
| 30 | 112 | 2.7 | 1128 | 2 T08312 | origin of replicat |
| 31 | 111.5 | 2.7 | 1902 | 2 B44858 | serine proteinase |
| 32 | 111 | 2.6 | 753 | 2 D72660 | probable aldehyde |
| 33 | 108.5 | 2.6 | 1167 | 1 A35066 | streptococcal C5a |
| 34 | 108.5 | 2.6 | 1557 | 2 T13160 | protein CNK - fruit |
| 35 | 107.5 | 2.6 | 1902 | 2 B45764 | serine proteinase |

| | | | | | |
|----|-------|-----|------|----------|---------------------|
| 35 | 107.5 | 2.6 | 695 | 2 D71283 | probable translati |
| 37 | 107 | 2.5 | 679 | 2 C71413 | hypothetical prote |
| 38 | 106.5 | 2.5 | 1902 | 2 S06997 | serine proteinase |
| 39 | 106.5 | 2.5 | 444 | 2 T01721 | hypothetical prote |
| 40 | 105.5 | 2.5 | 1257 | 1 A41060 | neural cell adhesi |
| 41 | 104.5 | 2.5 | 701 | 2 S35430 | gag polyprotein - |
| 42 | 104.5 | 2.5 | 2475 | 2 T00047 | gelatin lyase (EC 4 |
| 43 | 104.5 | 2.5 | 1075 | 2 A57377 | transcription fact |
| 44 | 104 | 2.5 | 1330 | 2 B70836 | hypothetical prote |
| 45 | 104 | 2.5 | 1386 | 2 S73401 | MG064 homolog R02_ |

ALIGNMENTS

RESULT 1
A48592
transferrin receptor protein - Chinese hamster
C:Species: Citicellus griseus (Chinese hamster)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 20-Aug-1999
C:Accession: A48592
R:COLLAW, J.F.; Lai, A.; Domingo, D.; Fitch, M.; Hatton, S.; Trowbridge, I.S.
J. Biol. Chem. 268, 21686-21692, 1993
A:Title: YTRF is the conserved internalization signal of the transferrin receptor, an
A:Reference number: A48592; M01D:94012749
A:Accession: A48592
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-757 <COL>
A:Cross-references: GB:L19142; NID:g304528; PIDN:AAA0576.1; PID:g304529
C:Superfamily: transferrin receptor
C:Keywords: endocytosis; glycoprotein; transmembrane protein

Query Match 35.8% Score 1504.5: DB 2: Length 757;
Best Local Similarity 42.8% Pred. No. 1.5e-100;
Matches 334: Conservative 119; Mismatches 248; Indels 79; Gaps 19;

| | | | |
|----|-----|--------------------------------------------------------------|-----|
| QY | 35 | EEEEDEGEAEALAHFCPELRGPEPLSRPNLIPMAAGRAAPYLVTALIFT | 94 |
| DB | 44 | EEENTDNMKASVVRK---RLNGRLRCGT-----IAVVEFL | 78 |
| QY | 95 | GAPLLGYAF--RGSCQAC-----GDSVLVSEEDVNYPPDLDFHOGRLYWSDLQMF | 144 |
| DB | 79 | IGFMIGYGYCKRTEOKCVRLAETETGSEIIIOENIP-----QSSRLYMDLKKLL | 131 |
| QY | 145 | LOFLGEGRLDTIROTSIREVRVAGSAGMAALTOIRALSNOKIDHWTDTHVYGLQFPD | 204 |
| DB | 132 | SEKIDALEFTDTIKQLSOTSRASQKDEMLAYIENOFDFKLSKVRDEHYVAKIOVG | 191 |
| QY | 205 | PAHPTLTHVDKAGVGEOLPLEDDPVYCPYSALGNVGEIVVAHYGPEDLODFRANGV | 264 |
| DB | 192 | SAQNAVAVIIVN---GSDLVENPGVAVASKATTVSGKLIHNFGRKKFEDLKY--- | 246 |
| QY | 265 | DPV-GRLLIVRVAIYSEAKVTNAODEGAGGVLIYEPADFSODPPKSLSSQAAVGHV | 323 |
| DB | 246 | PVNGSLIVYRAGKITFEKVAANAQSFNAIGVLIYMDQTKF-----PVYAEALSLFEGA | 298 |
| QY | 324 | HLGSDPPTPGFPSPFNQOPFPVASSGLPIPAQISADINSLRLKAKGVAAQDEMGGS | 383 |
| DB | 299 | HLGSDPPTPGFPSPFNHQPSPSSGSLPIPVDTISKRAEKLQANNE-TNCPSPWTD | 357 |
| QY | 384 | LLGSPYHLGPPRLRLVNNHRTSTPINNIGCIEGRSEPPHYVYVIGAQDAMGGAAS | 443 |
| DB | 358 | SL-CKLESSGGINLVSNVNLKERTIRINITGVKGFEPFRIVYVIGQORAMGGAAS | 416 |
| QY | 444 | AVGATLLELVRTSSWVS-NGFRRRSLLFISWDGDFGSGVSTEWLEGLVSLHLKAV | 502 |
| DB | 417 | SVGSLILKLAKAQSADVWSRGCFKPRSRIIFASWSAGDFGAVGATWLEGLYSHLKAF | 476 |
| QY | 503 | VYVSLDNAVIGLDDKDFHAKTSPLTSLTESVLYKQVDSPRHSQOTLYEQVVFNPMSDAVI | 562 |
| DB | 477 | TYINDKVVLTGRNKFVASPLLYTLLETKTMDVDRHP-IDSKPLRYR-----DSNWKSV- | 530 |

QY 563 RPLPMSSAATFAFYGVAVESFEMDDQAYFFLTKEPTYENL-HKYUGRLPAVAQA 621
D 530 EDLSLNAEAPFLAYSGIPAVSWFCENED-YEYLDITNLDITYEKLIOKVPQ-LNKKVRA 586
QY 622 VAOLAGOLRLRLSHDRLLPLDFGRYGDVVRHIGNINEFSGDKARGLTLQWYASRGDY 681
D 587 AAEVAQOFITKRLHDIELNDYDMYNNKILSFYKELNQFADIKAGLSIQWLYSARGDF 646
QY 682 IRAAEKLRQETYSSEDERLTPMYNVRIMRVEFFYLSQYVSPADSEFRHIFMGSDHTL 741
D 647 FRATSLTDTDFHNAEKTRNFVREINNRIKMYEHFSLPYVSPRESEFRHIFWGSCHTL 706
QY 742 GALLDHLRLRSNMSGTPGATSTGQESRRRQALLTTTGAAANALSGDVNINDNF 801
D 707 TALVENLKLKQNNSS-----AFNETLFRNQLALATWTIGQVANALSGDITWDIDNEF 757
RESULT 2
JXHU
transferrin receptor - human
N:Alternate names: CD7L; P90
M:Contains: 85k serum transferrin receptor
C:Species: Homo sapiens (man)
C:Date: 25-Feb-1985 #sequence.revision 25-Feb-1985 #ext.change 22-Jun-1999
R:Schneider, C.; Owen, M.D.; Banville, D.; Williams, J.G.
Nature 311, 675-678, 1984
A:Title: Primary structure of human transferrin receptor deduced from the mRNA sequence.
A:Reference number: A93343; M0ID:85012743
A:Accession: A93343
A:Molecule type: mRNA
A:Residues: 1-760 <SCH>
A:Cross-references: GB:U01060; NID:q37432; PIDN:CAA25527.1; PID:q37433
R:McClelland, A.; Kuhn, L.C.; Ruddle, F.H.
Cell 39, 267-274, 1984
A:Title: The human transferrin receptor gene: genomic organization, and the complete pri
A:Reference number: A90856; M0ID:85048936
A:Accession: A90856
A:Molecule type: mRNA
A:Residues: 1-760 <MC>
A:Cross-references: GB:M1507; NID:q39915; PIDN:AA61153.1; PID:q39916
R:Shih, Y.J.; Baynes, R.D.; Hudson, B.G.; Flowers, C.H.; Skikne, B.S.; Cook, J.D.
J. Biol. Chem. 265, 19077-19081, 1990
A:Title: Serum transferrin receptor is a truncated form of tissue receptor.
A:Reference number: A36597; M0ID:91035436
A:Accession: A36597
A:Molecule type: protein
A:Residues: 101-103, 'X', 105-108, 'X', 110-119 <SHI>
A:Experimental source: serum
R:Gopollno, M.; Migliorini, M.; Argaves, W.S.; Dedhar, S.
Biochem. J. 306, 129-134, 1995
A:Title: Identification of a novel form of the alpha(3) integrin subunit: covalent assoc
A:Reference number: S54327; M0ID:95169043
A:Accession: S54327
A:Molecule type: protein
A:Residues: 288-302;694-708;721-730 <COP>
R:Alvarez, E.; Girones, N.; Davis, R.J.
Biochem. J. 267, 31-35, 1990
A:Title: A point mutation in the cytoplasmic domain of the transferrin receptor inhibits
A:Reference number: S09039; M0ID:90226333
A:Accession: S09039
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19, 'C', 21-61 <ALV>
A:Note: mutant defective in endocytosis
C:Comment: This transmembrane glycoprotein exists as a dimer of similar or identical cha
ity acyl groups. The amino end of each chain lies within the cytoplasm and a stop-transf
embrane, may also serve as an internal signal sequence.
C:Comment: The expression of this receptor, involved in the regulation of cell growth, c
C:Genetics:
A:Gene: GDB:TFRC
A:Cross-references: GDB:120433; OMIM:190010

A:Map position: 3q26.2-3q26.2
C:Function:
A:Description: mediates cell iron uptake by binding, internalizing, and recycling the
C:Superfamily: transferrin receptor
C:Keywords: glycoprotein; iron transport; receptor; transmembrane protein
F:1-57/Domain: intracellular #status predicted <INT>
F:20-24/Region: tyrosine-based endosomal/lysosomal sorting signal
F:58-61/Region: stop-transfer sequence
F:62-89/Domain: transmembrane #status predicted <TM>
F:89-760/Domain: extracellular #status predicted <EXT>
F:101-760/Product: 85k serum transferrin receptor #status predicted <MAP>
F:251,317,727/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 35.6%; Score 1496.5; DB 1; Length 760;
Best Local Similarity 44.5%; Pred. No. 5,7e-100;
Matches 327; Conservative 119; Mismatches 230; Indels 59; Gaps 20;

QY 84 YLVLTALLFTGAFLLGYAF-----RSCQAGDSVLVSEEDVYEPDLDPEHQS-RLY 136
D 68 YGTIAVVEFFLGFMTGYLGCKGVKTECEPLACTESPVE---EPGEFPAARRLY 123
QY 137 WSDLAQMFLOFLGEGRLDPTIR---QTSLERYAGSAGMAALTODIRALSRKLDHWT 193
D 124 WDLKRLKLEKLDSTDTFTIKLNSVYPRHAGSQKXENALVYENQFREKLSKVMR 183
QY 194 DTHYVGLQPPDPAHPNTLWVDEAGKVGSQLPLEDDVYCPYSAIGNVTGELYAHYGRP 253
D 184 DQHFVLIQVKSQ--NSVIYIVDKNGRL--VYLVENPGYVAASKATVGVKLVHANFGRK 240
QY 254 EDLQDILRANGVPV--GRLLVRYGVISFAOKYTNADDFAGOGVLIPEPADSQDPKPS 312
D 241 KDEEDL-----YFVNGSIVYRAGKITFEAKVANAALNAGVLIYMDTKF-----PI 290
QY 313 LSSQAVYGVHVLGCTDPTTPGPFSPNQFPFVASSGGLPSIPAPISADIASRLRLK 372
D 291 VNAELFEFGAHLGCTDPTTPGPFSPFNHQFPSPSSGLPNIPYQISAAAEKLEFGN 350
QY 373 GPVAPQEQSGSLGSPYHL--GGPRLRLVYNNHRTSTPINNIFGCIERSEPEPDHYVIG 430
D 351 GD-CPSPDK---TDSYCRVYTSKKNVKTLSVNLKEIKLINFYIKGFVEPDHYVVG 406
QY 431 AORDANGPGAAASVATLILLEVRESSM--SNGRPRLSLFISWMDGSGVSGTEM 489
D 407 AQRDANGPGAAASVGTALLKLAQMFSDVYLKDGQPSRSITFASWSAGDGSVATEM 466
QY 490 LGEYLSVLHLKAVVYVSLDNVAVLGDDKFAKTSPLTSLIESVLYKQVDSPNHSGQTLYE 549
D 467 LGEYLSVLHLKAVTYINLDKAVLGTSNFKVSAAPLYTILEKTMQVKKP-VYGFLLY 525
QY 550 VYFTNSWDAEVYRPLPMSSAATFAFYGVAVESFEMDDQAYFFLTKEPTYENL 609
D 525 ---DSNM-ASKVEKLTLDNAEPFLAYSGIPAVSWFCFEDD-YEYLDITNLDITYELE 578
QY 610 VLQGRP---AQAQVLAQGLRLRLSHDRLLPLDFGRYGDVVRHIGNINEFSGDLK 666
D 579 ----RIPELNAKARAALAEVAGQVILTHDVELNDYENKQSLSEFYVDLNDYRADIK 634
QY 667 RGLTLOWYSARGDYIRAAEKLRQETYSSEDERLTPMYNVRIMRVEFFYLSQYVSPAD 726
D 635 MGLSLQWLYSARGDFRATSRLLTDTGNAEKTRNFVREINNRIKMYEHFSLPYVSPRE 694
QY 727 SPFRHIFMGSDHTLQALLDHLRLRSNMSGTPGATSTGQESRRRQALLTTTLOGA 786
D 695 SPFRHIFMGSGHTLTPALLLENILKRNNG-----AFNETLFRNQLALATWTIOGA 745
QY 787 ANALSGDVNINDNF 801
D 746 ANALSGDVNINDNEF 760

RESULT 3
S29548


```

QY 65 RPRQNPILPWAAGRRAPYLVLTALLIFGAFLLGYVAFRGSOAGDSVLVYSEDVNY 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17 RPR-----WLCAG-----ALVLGAFLLGFL-----FGWFKSSNEVTNT 52
QY 125 EPDLDFHOGRLWYSDLOAMFLOFEGEGLDITROTSLRERVAGSAGMAALTODIRAL 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53 TPK---HNKMAFIDELKA-----ENIKKFLYNTQIPIHLAGTEQNFOLAOIQSQMK 101
QY 185 RQGLDHVWIDTHVGLQFPFPAHPNTLHWYDEKGYEQPLPDDP-----DYCY 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102 EFGDLSEVLAHYDVLSTYPRKTHPNYISINEDGNEIFNTSLEFPPEGYENSDIYPP 161
QY 236 SAI---GNVYGEVLAHYGPEDLQDL-RARGVDPVGRLLLVRYGVSFGOKYTNADPG 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 162 SASSPQGMPEGDLVYNYATFEDFFKLERDKMKNCSKTIYARIGKVRGNKKNAQLAG 221
QY 292 AOGVLVYPEPADFSODPKPSLSQOAVYGVHL-----GTGDPYTPGPFSENOTQFP 344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 222 AKGVILYSDPADYFAPGVKYPDGWNLPGGVGQNGNTLNNGAGDPLTPGIPANEVYRR 281
QY 345 PVASS-GLPSIPAPISADIASRLKLPVAP-QEWGSSLGSPYHLPG-----P 395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 282 GIEAAGLPSPYHPHPIGYDAOKLEKMGSGAPPDSWRS-LKVPYVPGFTGNFSTQ 340
QY 396 RLRLVNNHRTSTPINNIFGCIERSRSEPDHYVYVIGAQRDAMGCAANSVGTALLLEVR 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 KYAMHISTNEVTRITNYIGTLRGAVEPRYVILGHRDSWVFEGIDPQSGAAVHEIVR 400
QY 456 TFSMSYNGFRRRLSFLFISMDGDFGVSSTWLEGLYSLVHLKAVVYSLDNALVGD 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 SFTLKEGRRPRRTILFAWDAEEFGLGSTMAEENSLLQERGAYINADSSIGNY 460
QY 516 KEFAKSPILTSLESYLKQVDSPNH--SCQTYEYVFTNPMSDAEVIKPLPMDSSAYS 573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 TLRVDSPTLWYSVHNLTKEKSPDEGFEKGSYTESWTKSPSEFGSMRISILGSGND 520
QY 574 FTAFC--VGPAVEFSMEDD-----APPLTKEDTYENLKHVLOGRLPAVAQAQOL 625
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 521 FEVFFORLGSARGARTKMTNKSFGYPLYSVETYLEVEFYDPMF-KYHLTAAY 579
QY 626 AGOLLRLSHDRLLPLDFGRYGDVVLRIHGNLNEFS---GDLKARGLTLOWYASAGDY 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 580 RGMVFEIANSIYLPFCRCRYAVVLRKYADKIYSISKHGOEKMTYVSVDLSFSAKNF 639
QY 682 IRAAEKLRQEIYSEERDEKLTMYNVRIMRVEFYLSQYVSPADSEF-RHIEGRGDAT 740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 640 TEIASFSEERLQDPDKSNPIVLRMMNDQLMFLERAFIDPLGLP-DRFPYHVYAPSSHN 698
QY 741 LGALLDHLRLRNSNSTPGATSSTGQESR-----FRQLALLTTLQGAANALS 791
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 699 KYA-----GESFPIYDALIEDIESKVDPSKMGVEKROIYVAFTYVAAAETLS 747

```

Best Local Similarity 25.6%; Pred. No. 1,3e-23;
Matches 182; Conservative 109; Mismatches 254; Indels 167; Gaps 32;

```

QY 24 QRVGPKRKHLE-----EEDGEGAEFTLHFCPMELRGPE-----PLGSPR- 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 ERKREKEGQLKERTVYNNADPDNEAEATGLQYSETTRDNEBSMNSFLTISNRG 66
QY 68 QPNILPWAAG-----RRAPIYLVLTALLIFGAFLLGYVAFRGSOAGDSVLVY 118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 RSNYISSIVSGYEIMKEHMKEFMYLILASLLMYG-FVAAF-APRTS-----L 114
QY 119 SEDVNEPDLDFHOGRLWYSDLOAMFLOFEGEGLDITROTSLRERVAGSAG-MAALTQ 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 SRDPR-----RHSRRLTNAEYRITLNSLQO-----ENRAKEHYVYAGTMSGAS 161
QY 178 DIRAALSROKLDHVTWIDTHY---VGLQFP---DPAHPNTLHWYDEKGYEQPLPDDPV 231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 162 D--SFTFKYTLDE-FLDMGKPKRVEKYPYIGPVDITNAPL--ENKQVYVEASMLEDRV 216
QY 232 -----YCPYSAIGNVTGELYAHYGRPEDLQDLRARGVDPVGRLLLVRYGV 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 217 KQDPASHARRKRGKGFHOYSKNGSVTARVYFCNGYSISDYKLLKKNIDIEDKIHVRSGR 276
QY 278 ISFAOKYTNADFGAGGVLYPEPAD-----FSQDPPK-PSLSQOAVYGVH 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 277 ILPLGLKYNALYGAASSVILYTDPPDGKYTEENGELHYGPARNPSTYLRDSV-NYFS 335
QY 325 LGTDPYTPGFPSP--FNQTFPPVASSGLPSIPAPISADIASRLKLPVAPQEMOG 382
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 DTPGDPTTPPYPKSDSTHEMSPVGR--VPRISVPMASADVPILERLN----- 384
QY 383 SLIGSPYHLGPPRL-----RLVNNHRT-----STPINNIFGCIEG 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 ---GRGFIQPGSINIKDFGFTGSSSIDKYNHNLTYNIKEMSVYEVISPGIF--TEG 438
QY 420 RSEPDHYVYVIGAQRDAMGPAAKSA-VGTALLLEIVYTFSSMSYNGFRRRLSFLTMDG 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 439 E-----ITIGHRDSIASSAGDANGSAILLEIAKMSKLLKHGKPLPKLISMDG 492
QY 479 GDFGVSSTWLEGLYSLVHLKAVVYSLDNALVGDKFAKTSPLTSLIESYLKQVDS 538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 493 ERSGLLGSTIYAEAAHAILRRRLVYINDNALISGNT-FICKANPLLDVITYEAKLTFF 551
QY 539 PNHSGQTYEQVVYFNPMSDAEVIKPLPMDSSAYSFTAFYGVPAVEFSMEDDQAYFLH 598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 552 NGHEMDSLFDMKRYTSNA---TISLIDGLSSTYSFQYHGVPAHFOFANNTSGAYVH 607
QY 599 TKE-----DHYEN---LHKVLOGRLPAVAQAQOLGOLLIRSHRL 638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 608 SNSVFDSPVWLEKFTNSDYKLNHTM-----AMFVGLTTLMLSENEL 648

```

RESULT 7
57149
probable membrane protein YJR126c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein J2050
C:Species: Saccharomyces cerevisiae
C>Date: 23-Aug-1995 #sequence_revision 08-Sep-1995 #text_change 14-Nov-1997
C:Accession: S57149
R:Rose, M.; Koetter, P.; Entian, K.D.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56848
A:Accession: S57149
A:Molecule type: DNA
A:Residues: 1-811 <ROS>
A:Cross-references: EMBL:249626; NID:g1015854; PID:g1015855; MIPS:YJR126c
C:Genetics:
A:Map position: 10R
C:Keywords: transmembrane protein

Query Match 10.4%; Score 438.5; DB 2; Length 811;

RESULT 8
565188
secretory protein SSP134 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein P2267; protein YPL176c
C:Species: Saccharomyces cerevisiae
C>Date: 10-Dec-1994 #sequence_revision 31-May-1996 #text_change 24-Sep-1999
C:Accession: S65188; JH0485
R:Rees, V.; Rechmann, S.; Nentwich, U.; Voss, H.; Ansoorge, W.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S65183
A:Accession: S65188
A:Molecule type: DNA
A:Residues: 1-783 <BEN>
A:Cross-references: EMBL:273532; NID:g1370370; PIDN:CAA97883.1; PID:e246908; PID:g137
A:Experimental source: Strain S288C (AB972)
R:Sidhu, R.S.; Mathewes, S.; Bollen, A.P.
Gene 107, 111-118, 1991
A:Title: Selection of secretory protein-encoding genes by fusion with PHOS in Sacchar
A:Reference number: JH0483; M0ID:92077420
A:Accession: JH0485
A:Molecule type: DNA

A:Residues: 66-149, 'P', 151-318, 'R', 320-414 <SUD>

A>Note: the authors claim that sequence of residues 41-57 is signal sequence

A>Note: the authors translated the codon TTA for residue 34 as Phe, GAT for residue 201

C:Genetics:

A:Gene: SSP134

A:Map position: 16L

C:Superfamily: secretory protein SSP134

C:Keywords: glycoprotein

F:139,213/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.1%; Score 297.5; DB 2; Length 783;

Best Local Similarity 22.4%; Pred. No. 1.9e-13;

Matches 126; Conservative 107; Mismatches 256; Indels 71; Gaps 19;

```

OY 152 RLEDTIRQTSLEKRVAGSAGMALTDIIRALSRKLDHWTDTHYVGLQPPDAPHPMTL 211
DB 160 KIERPLEYISSMPSHSGSDAIAIRHYIKESFDKNGIRLAGEEFEMAYSNYGNV---SL 216
OY 212 HNVDSAGVGEQLEPDDPVYCPYSAIGNVTG-ELVYAHYGPEPDQLDARGVDPVGR 270
DB 217 RYVSKDTEGEDIPLINEEN-FNPMHNGQLNIPVIYAKKASLDWMASQDGLNGDFI 275
OY 271 LLVRGVISFAQKVTNAODFGAGVLIYEPADFSQDPKPSLSSQQAAYGVHGLGTGP 330
DB 276 LLVHGDIYFQGMLT-ADEYGAKAIIIFSEPYQDNKD-----VIQMSVALPQYGTGDA 328
OY 331 YTPGPFENQTOFPYVASSGLSPISAPQISADIASRLKRLKG---PVAQEMQSLGGS 387
DB 339 LTPWEGSIRDPIDATEKACLPKIPISANQCKILAILSDTGKVFESNNLFSGSLNDC 388
OY 388 PYHLPGPRLRLVNNH-RTSPINNIEGCIERSPEPDHYVYIGARODMGFGAKSAV 446
DB 389 -----RLDLVOTAIHERPHVDIVKIEGSEQAGRAIYIAPRNASVGTMYPSFG 440
OY 447 TAILLELVTFSSAYSN-GEPRRSLLFIEMDGDGSGVSTEMEGYSLVHLKAVVY 505
DB 441 TVYLLSLIOYGEVYKFKPMKPLRNITFISFGSEFENEGALIELMEKREDA--LKSEIYT 498
OY 506 SLDNAVLG--DD--KFHAKTSPLLTSLIESVLK-----QVDSPNHSGOTLYEQVVF 555
DB 499 ILIDVQIGIEMDSDNMLEICHPLELDLFGKNMSTRKFNKVNVAHQFG----- 547
OY 556 SMDAIVIRPLMDSSAYSTAFVGVPAVEFSEMEDQAPFLHTKEDITYENLHKYLGRL 615
DB 547 DMTPLYAGIIPV-----AITSSGVV-----NREHP-IYTVEDKDFDKLRODK 591
OY 616 PA-VAQAVALAGOLLIRLSHDRLPLDGRYGDVYLRIHGNLFSGDLKARGLTLQWY 674
DB 592 KGEVSEIMLYEVEKSLIEDDPFIPFISNVVDPLSTLKLQKRECP-----TVNPD 645
OY 675 YSARDYIRAAEKLROEIIYSSSE 696
DB 646 EVELGTITMENTKIQEKKWSE 667

```

RESULT 9

567153 Probable membrane protein YOR256c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein O5330

C:Species: *Saccharomyces cerevisiae*

C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 24-Sep-1999

C:Accession: S67153

R:Jauniaux, J.C.; Poirey, R.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67143

A:Accession: S67153

A:Molecule type: DNA

A:Residues: 1-609 <JAN>

A:Cross-references: EMBL:275164, NID:g1420578, PIDN:CAA99478.1, PID:e252109, PID:g142057

A:Experimental source: strain S288C

C:Genetics:

A:Map position: 15R

C:Superfamily: secretory protein SSP134

C:Keywords: transmembrane protein

F:126-142/Domain: transmembrane #status predicted <TM>

Query Match 5.8%; Score 243; DB 2; Length 809;

Best Local Similarity 21.1%; Pred. No. 1.7e-09;

Matches 169; Conservative 103; Mismatches 299; Indels 230; Gaps 30;

```

OY 83 PYLVLTALLIFTGAFLLGYAERG---SCQACGDSVLVYSEDEVNE---PDIDFHQGR 136
DB 120 PEILRRFFYIIFMSPIAYVYVSSGLFNEKASGSGMFSQDILFEVAKKASVDL----- 174
OY 137 WSDLOAMFLQELGSRLEDITRQTSLEKRVAGSAGMALTDIIRALSRKLDHWTDTH 196
DB 174 -----AKFERDLEYISSMPSHSGSDAIAIRHYIKESFDKNGIRLAGEEFEMAY 219
OY 197 YVGLQPPDAPHPMTLHWVDEAGKVEQLPLE-DPVYCPYSAIGNVTG-ELVYAHYGRPE 254
DB 220 SVYSNYPGNV---SISTYDNK---NEKHDLSEKENENFPLSKVSLIYGGKTTY 273
OY 255 DLQDLRARGVDPVGR---LLVRGVISFAQKVTNAODFGAGVLIYEP---ADFSQD 307
DB 274 DLQHLKDSKTIEDGKDYVLLQYDKLVS--QQVLAIEKFGAKAVIIFISEPYGENIDV 331
OY 308 PPKPSLSSQQAAYGVHGLGTGDPPTPGSPFNQTOFPYVASSGL----- 352
DB 332 KP-----VGLPYQSTGD-----ASGLNMDGSPVEEKDHKFWR 363
OY 352 -PSIPAPISADIASRLRLK-KGPVAPOEMQSGSLGSPYHLGPPRLRLVNN-----NH 404
DB 364 OTHIPTITISIRQGEKELLSRSLSSGVTYD-----GMSDSSNSGKMDVILIDVLO 417
OY 405 FTSPINNIFCIEGRSEPDHYVYIGARODAMGPAKASAVGTAILLELVTFSSM-VSN 463
DB 418 REKHPIPTIVKIEGREGQSKAIIIAASRNSINGTYPNFGTALLISTVOLFQEVKKYK 477
OY 464 GFRPRRSLLFIEMDGDGSGVSTEMEGYSLVHLKAVVYSLDNAVLG---DDKF-HA 519
DB 478 GWKPLRNITFISFGTEFNVAAGSELVEQRLLT--PLKEIYSLIDISQIGIPFAEKYENG 535
OY 520 KTSPLLTSLISVLKQ-----VDSPNHSGOTLYEQVVFTRPSMDAIVIRPLP 566
DB 536 KTRGSLSEIETPLPLKFFNRNAHGFDSVQNVQHYG-----DMTPPLANIP 583
OY 567 MDSSAYSTAFVGVPAVEFSEMEDQAPFLHTKEDITYENLHKYLGRLPAVAQVAOLA 626
DB 584 VSVISDSSTRNDPPT-----ETSPDKFERVEKILEDE--QNOQSVKDLL 626
OY 627 GQLL---IRLSHDRLPLDGRYGDVYLRIHGNLFSGDLKARGLTLQWYYSARGDYIR 683
DB 627 VYLLHISMLIDDPDLHFDIISY----- 650
OY 684 AAELKROEIIYSSSEDERLTRYNYRIMVEFEYL-----SOYVSPADSPFRI 732
DB 650 -----VEDIDERLQLEQAYPEKLNFTSIKGLLEFWKKGSFMAWSVQG-WENI 697
OY 733 FWGRGD-----HTLGALLDLRLRSNNGTGPAT-----SSTGFQESRFR 774
DB 698 VMSHGDGIEPSSLINRTWKKLINIGRRCSPAGLPNRSFYKNVLGPTLIQDKSKN 757
OY 775 QLALLTWLQGAANALSDVW 795
DB 758 GGNVDEWTFPGVMDAIYDDW 778

```

RESULT 10

568317 transferrin receptor - guinea pig (fragment)

C:Species: *Cavia porcellus* (guinea pig)

C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 20-Aug-1999

C:Accession: S68317

R:Gosielska, A.; Mahmoodian, F.; Peterkofsky, B.

344 PVV-----ASSGLPSIPAPISADIAS-----RLLRKLKGPVA-----PQEQW---GSL 385
108 PPIIOQAVTATLPOVDASQITISLSLESTFNRTTSSGQASDWMASEQWALSASL- 167
386 GSPYHAGPRLRLVNNHRTSTPINNIFGCTEGRSEPDHYVYVIGADAW----- 437
167 -----FNASVKQVSHSGYNO---KSVYMTITGSEAPDEWLVIGHLDTGSHTNQOS 216
437 -GPGAKASAVGTAILLLELVRTFSSVMSNGFRRLSLFIISMDGGFGVSGSTEMLEGLS 495
217 VAPGADDASGIAAVTEVIRVLS---ENNFOKRSIAFMAYAAEEVGLRGSODLANOY-- 272
496 VLHLKAVVYVSLDNVLDGDKFHAKTSPULTSLIESVYKQVDPSPHSGOTLEQVYFNP 555
272 -----KSEGN-----VVSAL-QDLMYKQSA--QDVVFITD 301
556 SWDAEVRPLP-----MDSAYSEPTAF-----GVPAV---EFSFMDDOAY 594
302 YTDNFTQYLNQLMDEYLPSLTYGFDTCGYACSDHSMHNGYPRAMPFESKFNQYN--- 359
595 PFLHTKEDTYEN 606
359 PRIHTTODTLAN 370

RESULT 13
F75191
hypothetical protein PAB0027 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: F75191
R:Anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: F75191
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <KAW>
A:Cross-references: GB:A1248283; GB:A1096836; NID:g5457433; PIDN:CA848981.1; PID:e151487
C:Genetics:
A:Experimental source: strain Orsay
A:Gene: PAB0027

Query Match 3.2%; Score 136; DB 2; Length 561;
Best Local Similarity 20.3%; Pred. No. 0.049;
Matches 128; Conservative 93; Mismatches 236; Indels 172; Gaps 31;

147 FLGGRLEDITIGTSLREYVAGSAGMALTDIDIRALSRQKLDHWMTDTHYGLQPPDA 206
15 FKSENVFRDYY-EISRYHRIQGSREIVDAEYVAKRLSIGLEYFLNDRYGKRY---- 70
207 H---PNTLHWYDEAGVGEOLPLED-----PDYCYSAIGWYVGEVYAHYGRPED 255
70 HLTLSPIAMELKRK-----LEFEDKVLTTNSPLLMAHSSGAEGLVLF--KDED 123
256 LODLRARGVDPGRLLLVYRVGVISFAQKYTNADFGAQGVLLYPPADFSQDPKPSLSS 315
124 WEKAE-----GKIVLGE--DWREAVKRNAGARAFIAYRK----- 159
316 QQAVYGHVHLGTG--PYTPGFSPNOTOPFPVASSGLPSIPAQIISADIASRLRLKLG 373
159 -----GTGAPFPYIGLFTKKDLEW-----ARIPALVPETFANELINKAK- 200
374 PYAPOEMOGSLIGSPYHLGPGRLRLVNNHRTSTPINNIFGCTEGRSEPDHYVYVIGAQ 433
200 -----KG-----GAKVVEV---ETELKESVDLVYLAKEIPYLLFSAH 238
434 DAMGGAASAVGTAILLLELVRTFSSVMSNGFRRLSLFIISMDGGFGVSG-STEMLE 492
239 CHPRRGANDNAGSAMLEIARVLK-----GKEGRVGFALWIP 278

493 Y--LSVLHKA---VYVYSLDNVLDGDKFHAKTSPULTSLIESVYKQV 536
279 YHGTOAIPKRAKDEIYANILDMWGSSEDRKSTIMLVRTPLSRFSLPEVLTGTQOLI 338
537 DS--PNHSGQLYEQVVFETNPSMDAEVIRPLPMSSASVFAFVGPVAFERSFEDDOAY 594
339 NSGKSFSGSFL-----PRMKLKEY-PYEMGSDHDIFNIF-GIGVMPITMPS--- 386
595 PFLHTKEDTYENLHKVLOGRLPAVAQAVALAGOLLIRLSHDLPLDPFGRYGVDVLRHT 654
386 -YHTSADTPE--KLSTLESLEIIGKAAVAATA-VFLKAKEKEELEERVARG----FAKKYL 436
655 GNL-----NEFSGDLKARGLTLOWYYSARGDIFRAEK-----LRQIYSSE 696
437 GELMLERKIEVAESLVNMGSLFLGNGHIEEBGS-IKWEKGIISLKALEYKE 495
697 ERDERLTRMYNVRIMRY---EFYFLSQYV 722
496 ELAKRYKEITEERMIASHIHEYLMLSEIV 524

RESULT 14
E71223
hypothetical protein PH0048 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
C:Accession: E71223
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5: 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A:Reference number: A71000; MUID:98344137
A:Accession: E71223
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-565 <KAW>
A:Cross-references: GB:AP000001; NID:g3236128; PID:d1030059; PID:g3256433
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by Genba
C:Genetics:
A:Gene: PH0048

Query Match 3.1%; Score 131.5; DB 2; Length 565;
Best Local Similarity 19.3%; Pred. No. 0.1;
Matches 122; Conservative 106; Mismatches 226; Indels 177; Gaps 34;

146 QELGGRLEDITIGTSLREYVAGSAGMALTDIDIRALSRQKLDHWMTDTHYGLQPPDP 205
17 KFNENNVFRDYY-ELSKRHRIQGSRDYKATEYVASRLEELGIDYKFLNDIYDGKSY--- 73
206 AH---PNTLHWYDEAG---KVGEOLPLEDPPDYCYPSAIGNVYGEVYAHYGRPEDLOD 258
73 -HLTLPSPIGMEIVEGSLKFKEKELTNSPLLVMAHSPSGAEVGPVIF--REEDWEK 129
259 LRARGVDPVGRLLLVYRVGVISFAQKYTNADFGAQGVLLYPPADFSQDPKPSLSSQA 318
130 AD-----GKIVLGE--DWREAYRKANENHGAKAFLAYRK----- 162
319 VYGHVHLGTG--PYTPGFSPNOTOPFPVASSGLPSIPAQIISADIASRLRLKGPVA 376
162 -----GTGAPFPYIGLFTKKDLEW-----AKIPALVPEESVANELEISK- 203
377 PQEMOGSLIGSPYHLGPGRLRLVNNH--RTSTPINNIFGCTEGRSEPDHYVYVIGAQ 433
203 -----KG-----GVEVEVSKTEIKDRETLPL--LYAKI--GEPV-YLLFSAH 241
434 DAMGGAASAVGTAILLLELVRTFSSVMSNGFRRLSLFIISMDGGFGVSG-STEMLE-- 491
242 CHPRRGANDNAGSAMLEIARILK-----GNGRGIGFALWIP 281
491 -----EGYLSVLHKAVYVYSLDNVLDGDKFHAKTSPULTSLIESVYKQ 535


```

Db      282 YHGQAFIPVYNLNE-IYANINLDMWGSSDRNSTIMLVRTPLSRFSLVGVLEFLKE 340
QY      536 V--DSPHSSQOTLYEQVFFNPSPDAEYIRPLPMDSSAISTAVGVPAVEFSMEDDQA 593
Db      341 VNAGKFSGSPL-PSYKFEKPEYE-----MGSDHDFNFYSIPGVPIW-----PDRY 389
QY      594 YPFLHTEKEDYEENLHKVLOGRLPAVAQVLAQGLLIRLSHDLPLDPRGVDDVLRH 653
Db      390 Y---HTSADIPD---KLSRLTSLITGRAVY-ATGLFTAKAEKEEVERVARG---FAMKY 438
QY      654 IGNL-----NEESGDLKARGLTLLQWVVSARGDYIPAAEK---LRQETYS 695
Db      439 LGELSMQRKTEVAESLVNMGSRDSGFLGLNVGHELESNG-VWVKMKKGITISNGLEYKN 497
QY      696 EERDERL-----TRKYNVRIMRVEYFFLSQ 720
Db      498 ESTARRLKEIVEEDRMIVHLH--EYLMISE 526

RESULT 15
A:1341
microbial serine proteinase (EC 3.4.21.-), minor (vpr), precursor - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 17-Jul-1992 #sequence.revision 17-Jul-1992 #text.change 29-Sep-1999
C:Accession: A41341, B41341, S39700, D69730
R:Strom, A.; Rulo Jr., G.A.; Theriault, K.A.; Dwyer, M.; Wilson, S.W.; Pero, J.
J. Bacteriol. 173, 6889-6895, 1991
A:Title: Cloning and characterization of the gene for an additional extracellular serine
A:Reference number: A41341; MWID:92041574
A:Accession: A41341
A:Molecule type: DNA
A:Residues: 1806 <SIO>
A:Cross-references: GB:W76590; NID:g143819; PIDN:AAA22881.1; PID:g143820
A:Accession: B41341
A:Molecule type: protein
A:Residues: 161-195 <SL2>
R:Glasner, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu,
A.; Rapoport, G.; Danchin, A.
Mol. Microbiol. 10, 371-384, 1993
A:Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region fr
A:Reference number: S39655; MWID:95020537
A:Accession: S39700
A:Molecule type: DNA
A:Residues: 1806 <GIA>
A:Cross-references: EMBL:X73124; NID:g413923; PIDN:CAA51601.1; PID:g580871
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caidwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Galled
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holappel, S.; Hosono, S.; Hullo, M.F.
Koelter, P.; Koningsstein, G.; Kroh, S.; Kunano, M.; Kurita, K.; Lapidus, A.; Lardinois
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mueel
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schreier, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serot
Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MWID:98044033
A:Accession: D69730
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-806 <KUN>
A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PIDN:CAB15835.1; PID:e1186308;
A:Experimental source: strain 168
C:Comment: The amino terminal sequence of the mature protein and a molecular weight of c
C:Genetics:
A:Gene: vpr
A:Start codon: TTG
C:Superfamily: microbial serine proteinase vpr; subtilisin homology

```

C:Keywords: hydrolase; serine proteinase
F:1-28/Domain: signal sequence #status predicted <SIG>
F:28-160/Domain: propeptide #status predicted <PRO>
F:180-948/Domain: subtilisin homology #status atypical <SBT>

Query Match 3.1%; Score 130; DB 2; Length 806;
Best Local Similarity 21.8%; Pred. No. 0.24; Mismatches 163; Indels 148; Gaps 23;
Matches 101; Conservative 52;

```

QY      81 AAPLYVTALLIFGAFILGVAFRSGCOACGSVLVSEDEVNT-EPDL--DFHQGRLYW 137
Db      164 SAPYIGA-----NDAMDLYT-----GKGIYALIDGVEENHDDLKKNFCQYQYGY- 210
QY      138 SDIQAFLQFL-----GEGRLEDYIROTSLERVAGSA----- 171
Db      210 -----DEVNDYDPRKPTGDPREGAHDGHTVAGTVAANGTIKGVAPDATLLAYRV 261
QY      171 ---GMAALTDQIRALSRKLDHWITDTHYVG--LQPPDAHPTLLHWVDEAGKV----- 221
Db      262 LGPEGSGTTEWVIRAGVERAVODGADVNNLSLGNLNNPDMATSTALDMAMSEGVVAVTSN 321
QY      221 -----GEQLPLEDDVY-YCPYS-----AIGNVT 242
Db      322 GNSGPNQWTVGSPTCSREAISSVGTATQLPLNBYAVTFSGYSYSAKMYGKNKEDVYKALNKE 381
QY      243 GELVYAHYGRPEDLDLRAGVDPVGRLLLVYRGVVISFAQKVTNAQDFGAQGVLIYEP-P 301
Db      382 VELVEAGSIGCAKDFE-----GKDLTGKVAVVKRKSIAFVDKADNAKAGAIQVYVNNLS 436
QY      302 ADFEQDPP-----KPSLSQQAIVGHVHLGTDPTTPGFPSPNOQFPFVPAVSSGLPSI 354
Db      437 GEIEANVPGMSVPTIKSLDGEKLVSAKAGE-----TKTTEFLVYSKAL--- 483
QY      355 PAQPIASDIASRLRLKLGCPVAPQEMOGSLGSPYHLGPPRLRLVYNNHRTSTPINNIF 414
Db      483 -GEYV-ADFSR-----GPMV-DTW---MIKPDIAFPCVNIYSTIPTHDPHPYQ-Y 527
QY      415 GCIERGSEPDHYV-----VIGAQRDANGPAGANSAV-GTAILE 452
Db      528 GSKQGTSMASPHIAGAVAIKQAKPKMSVEQIKAIINMTAVTLK 571

```

Search completed: January 12, 2000, 22:33:58
Job time: 69 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: January 13, 2000, 01:07:03 ; Search time 11.62 Seconds
(without alignments)
2004.979 Million cell updates/sec

Title: US-09-358-755-1
Perfect score: 4203
Sequence: 1 MERLWGLFQRAQQLSPRSSQ.....TLOGAANALSGDVWNIDNMF 801

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database: SwissProt_38:*

Word size: 0

Number of hits that pass the threshold: 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|---------------------|
| 1 | 1496.5 | 35.6 | 760 | 1 TRSR_HUMAN | P02786 homo sapien |
| 2 | 725 | 17.2 | 750 | 1 PSM_HUMAN | O04609 homo sapien |
| 3 | 438.5 | 10.4 | 811 | 1 YJ96_YEAST | P47161 saccharomyc |
| 4 | 182 | 4.3 | 455 | 1 YWAD_BACSU | P35152 bacillus su |
| 5 | 143 | 3.4 | 504 | 1 AMPX_VIBPR | O01693 vibrio prot |
| 6 | 130 | 3.1 | 806 | 1 SUBV_BACSU | P29141 bacillus su |
| 7 | 123.5 | 2.9 | 467 | 1 VL2_BPV2 | P06457 bovine papl |
| 8 | 120 | 2.9 | 1448 | 1 TRC5_ECOLI | P27190 escherichia |
| 9 | 112.5 | 2.7 | 1805 | 1 NEST_RAT | P21263 rattus norv |
| 10 | 111.5 | 2.7 | 1902 | 1 P2P_LACPA | Q02470 lactobacilli |
| 11 | 108.5 | 2.6 | 1167 | 1 SCPA_STRPY | P15926 streptococc |
| 12 | 107.5 | 2.6 | 695 | 1 EFG2_TRBPA | O83748 treponema p |
| 13 | 107 | 2.5 | 887 | 1 UFO_HUMAN | P30530 homo sapien |
| 14 | 106.5 | 2.5 | 1902 | 1 P2P_LACLC | P16271 lactococcus |
| 15 | 106.5 | 2.5 | 1902 | 1 P2P_LACIA | P15293 lactococcus |
| 16 | 105.5 | 2.5 | 1257 | 1 CAML_HUMAN | P32004 homo sapien |
| 17 | 104 | 2.5 | 519 | 1 AMYH_SACFI | P26989 saccharomyc |
| 18 | 104 | 2.5 | 1386 | 1 Y064_MYCPN | P75613 mycoplasma |
| 19 | 103 | 2.5 | 638 | 1 PAN1_RAT | P21676 rattus norv |
| 20 | 103 | 2.5 | 930 | 1 PRGR_RABIT | P06186 oryctolagus |
| 21 | 103 | 2.5 | 1295 | 1 PURL_ECOLI | P35254 escherichia |
| 22 | 102 | 2.4 | 1319 | 1 ACSC_ACERY | P19450 acetobacter |
| 23 | 100.5 | 2.4 | 2241 | 1 TEGU_HCMVA | P16785 human cytom |
| 24 | 100 | 2.4 | 624 | 1 AMYG_ARYAD | P42042 axulla aden |
| 25 | 100 | 2.4 | 580 | 1 GAG_FIV | P10262 feline leuk |
| 26 | 100 | 2.4 | 917 | 1 SLAP_THERH | P35830 thermus aqu |
| 27 | 100 | 2.4 | 1170 | 1 YKD8_YEAST | P32862 saccharomyc |
| 28 | 99.5 | 2.4 | 3567 | 1 ERY2_SACER | O03132 saccharopol |
| 29 | 99 | 2.4 | 632 | 1 RCAC_FREDI | O01473 tremella d |
| 30 | 99 | 2.4 | 726 | 1 YANA_RHISN | P55577 rhizobium s |
| 31 | 98.5 | 2.3 | 744 | 1 CA18_RABIT | P14282 oryctolagus |
| 32 | 98.5 | 2.3 | 539 | 1 GAG_MLVF5 | P26807 friend muri |
| 33 | 98.5 | 2.3 | 1902 | 1 P3P_LACIC | P15292 lactococcus |
| 34 | 98 | 2.3 | 519 | 1 AMYG_SACFI | P08011 saccharomyc |
| 35 | 98 | 2.3 | 703 | 1 YAI1_RHISN | P55492 rhizobium s |
| 36 | 97.5 | 2.3 | 2594 | 1 TLES_DROYI | P20806 drosophila |
| 37 | 97.5 | 2.3 | 868 | 1 ACO2_SYNY3 | P74582 synechocyst |
| 38 | 97.5 | 2.3 | 1217 | 1 EGF_MOUSE | P01132 mus musculu |
| 39 | 97 | 2.3 | 944 | 1 GYRA_SYNY3 | P73077 synechocyst |

ALIGNMENTS

| | | | | | | |
|----|------|-----|------|---|------------|--------------------|
| 40 | 97 | 2.3 | 631 | 1 | PTBA_ERMCH | P26207 erwinia chr |
| 41 | 97 | 2.3 | 771 | 1 | TLE3_MOUSE | O08122 mus musculu |
| 42 | 96.5 | 2.3 | 701 | 1 | GAG_RSV | P03322 rous sarcom |
| 43 | 96 | 2.3 | 1666 | 1 | CLH_SCHPO | O10161 schizosacch |
| 44 | 96 | 2.3 | 707 | 1 | KPCT_MOUSE | O02111 mus musculu |
| 45 | 96 | 2.3 | 574 | 1 | MP12_RAT | P48966 rattus norv |

ALIGNMENTS

| | | | | | |
|--------|-----------------------------------------------------------------------|--------------------------------|-----------|------|---------|
| RESULT | 1 | TRSR_HUMAN | STANDARD: | PRT: | 760 AA. |
| AC | P02786; | 21-JUL-1986 (Rel. 01, Created) | | | |
| DT | 21-JUL-1986 (Rel. 01, Last sequence update) | | | | |
| DT | 15-JUL-1999 (Rel. 38, Last annotation update) | | | | |
| DE | TRANSFERIN RECEPTOR PROTEIN (TR) (ANTIGEN CD71) (T9) (P90). | | | | |
| CN | TRFC. | | | | |
| OS | Homo sapiens (Human). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | | | |
| CC | Eutheria; Primates; Catarrhini; Homniidae; Homo. | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE; 85012743. | | | | |
| RA | SCHNEIDER C., OWEN M.J., BANYTLE D., WILLIAMS J.G.; | | | | |
| RT | "Primary structure of human transferrin receptor deduced from the | | | | |
| RT | mRNA sequence."; | | | | |
| RL | Nature 311:675-678(1984). | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE; 85048936. | | | | |
| RA | MCLELLAND A., KUHN L.C., RUDDLE F.H.; | | | | |
| RT | "The human transferrin receptor gene: genomic organization, and the | | | | |
| RT | complete primary structure of the receptor deduced from a cDNA | | | | |
| RT | sequence."; | | | | |
| RL | Cell 39:267-274(1984). | | | | |
| RN | [3] | | | | |
| RP | SEQUENCE OF 101-119 (SERUM FORM). | | | | |
| RX | MEDLINE; 91035436. | | | | |
| RA | SHIH Y.J., BAYNS R.D., HUDSON B.G., FLOWERS C.H., SKIRNE B.S.; | | | | |
| RT | "Serum transferrin receptor is a truncated form of tissue receptor."; | | | | |
| RL | J. Biol. Chem. 265:19077-19081(1990). | | | | |
| RN | [4] | | | | |
| RP | PALMITOYLATION OF CYS-62. | | | | |
| RX | MEDLINE; 87218484. | | | | |
| RA | JING S., TROMBRIDGE I.S.; | | | | |
| RT | "Identification of the intermolecular disulfide bonds of the human | | | | |
| RT | transferrin receptor and its lipid-attachment site."; | | | | |
| RL | EMBO J. 6:327-331(1987). | | | | |
| RN | [5] | | | | |
| RP | MUTAGENESIS OF CYSTEINES INVOLVED IN INTERMOLECULAR BONDS. | | | | |
| RX | MEDLINE; 90005427. | | | | |
| RA | ALVAREZ E., GIRONES N., DAVIS R.J.; | | | | |
| RT | "Intermolecular disulfide bonds are not required for the expression | | | | |
| RT | of the dimeric state and functional activity of the transferrin | | | | |
| RT | receptor."; | | | | |
| RL | EMBO J. 8:2231-2240(1989). | | | | |
| RN | [6] | | | | |
| RP | INTERNALIZATION SEQUENCE, AND MUTAGENESIS OF TYR-20. | | | | |
| RX | MEDLINE; 90130619. | | | | |
| RA | JING S., SPENCER T., MILLER K., HOPKINS C., TROMBRIDGE I.S.; | | | | |
| RT | "Role of the human transferrin receptor cytoplasmic domain in | | | | |
| RT | endocytosis: localization of a specific signal sequence for | | | | |
| RT | internalization."; | | | | |
| RL | J. Cell Biol. 110:283-294(1990). | | | | |
| RN | [7] | | | | |
| RP | MUTAGENESIS. | | | | |
| RX | MEDLINE; 94012749. | | | | |
| RA | COLLAHN J.F., LAI A., DOMINGO D., FITCH M., HATTON S.; | | | | |
| RA | TROMBRIDGE I.S.; | | | | |


```

RESULT 2
PSM_HUMAN STANDARD: PRT: 750 AA.
ID PSM_HUMAN 00409.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PROSTATE-SPECIFIC MEMBRANE ANTIGEN (PSM).
GN PSM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 93113576.
RA ISRAELI R.S., POWELL C.T., FAIR W.R., HESTON W.D.;
RT "Molecular cloning of a complementary DNA encoding a
RT prostate-specific membrane antigen."
RL Cancer Res. 53:227-230(1993).
CC - FUNCTION: MAY INTERACT WITH TRANSFERRIN OR ANOTHER LIGAND AND
CC POSSIBLY FACILITATES METASTATIC SPREAD.
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC - TISSUE SPECIFICITY: PROSTATIC EPITHELIAL CELLS.
CC - SIMILARITY: TO THE TRANSFERRIN RECEPTOR PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M99487; AAA60209.1; .
DR MIM; 600934; .
KW Antigen; Transmembrane; Glycoprotein; Signal-anchor.
FT DOMAIN 1 19 CYTOPLASMIC (POTENTIAL).
FT TRANSNM 20 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 44 750 EXTRACELLULAR (POTENTIAL).
FT MOD_RES 1 1 BLOCKED.
FT CARBOHYD 76 76 POTENTIAL.
FT CARBOHYD 121 121 POTENTIAL.
FT CARBOHYD 140 140 POTENTIAL.
FT CARBOHYD 153 153 POTENTIAL.
FT CARBOHYD 195 195 POTENTIAL.
FT CARBOHYD 336 336 POTENTIAL.
FT CARBOHYD 459 459 POTENTIAL.
FT CARBOHYD 476 476 POTENTIAL.
FT CARBOHYD 638 638 POTENTIAL.
FT SIMILAR 418 567 TO TRANSFERRIN RECEPTOR.
FT CONFLICT 354 354 R -> K (IN AA SEQUENCE).
SQ SEQUENCE 750 AA; 84330 MW; 73B47D90 CRC32;

Query Match 17.2%; Score 725; DB 1; Length 750;
Best Local Similarity 27.3%; Pred. No. 5.5e-45;
Matches 213; Conservative 142; Mismatches 323; Indels 102; Gaps 22;

```

```

DB 162 SAFSPQMEPEGLVYVYARTEDFEFLERDMKINCSGIYARVKGVRGNKVNQAOLAG 221
QY AAGVLIYEPPEADFSODPPKPSLSSQAVYGVHL-----GTGPPYTPGFSPFQOTQFP 344
DB 222 AKGVILYSDPADYFAPGVKSPYDGNWLPDGGVQRGNILNLNAGADPLPGYANAYARR 281
QY PAAS-GLPSPAPQISADIASRLRLKLGVPAP-OEMOGSLGSPYHLGPG-----P 395
DB 282 GIAEAVGLPSPVYHIGYDDAKLLEKMGSGAPDPSNRGS-LKPYVNGPQFTNFSQ 340
QY RLRLVYNNHRTSTPIINNIFGIEGRSEPDHYVIGAQRDANGPAGAKSAVGTALLLEVR 455
DB 341 KKMHIISTNEVTRILYVIGTLRGAVEPDRVYILGHDMSWFGIDPQSGAAVYHEIYR 400
QY 456 TFSSWVSGFPPRRRLTLISWDGDFGSGVSTEMLEGLYSLVHLKAVVYSLDNAVLDGD 515
DB 401 SFGTLKKGWRRRRRIILPASMDAEFGILGSTWEMEENSRLIOERGVAIYNADSSIEGNY 460
QY 516 KFHAKTSPLLTSLIESVVKOYDSPNH--SGOTLYQVVFNPMSDAEYIRPLPMDSAYS 573
DB 461 TLRVDCPLMTSLVNLTKELKSPDEGEKSLVESWTKKSPSPFGMPRIKLGEND 520
QY 574 FTAFA--GVPAVESEFEMDDQ-----APPLTKEDTYENLHKVLOGRLPAVAQAVAL 625
DB 521 FEVFEQRLGASGRARVYKNETNKFSGYPLHSVYELVELVEKYDPMF-KYHLTVAQV 579
QY 626 AGQLIRLSDRLPLDFGRYGVVLRHIGLNEFS---GDLKARGLTLQWVYSGADY 681
DB 580 RCGWFEFLANSIVLPFDCRDYAVLRKYADKIYSISMHPDMKYSVSPDLSFAVKNF 639
QY 682 ITRAKLQEIYSSSEERERLRTMNVNIMVEFFLQYVSPADSP-RHIFMRGQHT 740
DB 640 TEIASKESERLDDFKSPVILRMNDQMLERAIIDPLGLP-DRPYRRIYVPSNN 698
QY 741 LGALDHLRLKSNSSGTPGATSTGFEQSR-----FRQALTLTWLQGANALS 791
DB 699 KVA-----GESPGIYDALFDLESKVDPSKANGEVKQIYVAFTQAALFELS 747

RESULT 3
Y96_YEAST STANDARD: PRT: 811 AA.
ID Y96_YEAST 47161.
AC P47161.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 92.0 KD PROTEIN IN RPS5-ZMS1 INTERGENIC REGION.
GN YJR126C OR J2050.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA ROSE M., KOETTER P., EMTIAN K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: TO MAMMALIAN TRANSFERRIN RECEPTOR PROTEIN AND
CC PROSTATE-SPECIFIC MEMBRANE ANTIGEN (PSM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49626; CAAB9657.1; .
KM Hypothetical protein.
SQ SEQUENCE 811 AA; 92017 MW; 4F94BBF1 CRC32;

Query Match 10.4%; Score 438.5; DB 1; Length 811;
Best Local Similarity 25.6%; Pred. No. 4.2e-24;

```


DB 364 SLTGGSSDHVPFHAGIDSANFTW-----GDP-----ETEEVEPMWHPEDSIEHISK 412

RESULT 5
ID AMPX_VIBPR STANDARD; PRT; 504 AA.

AC 001693;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE BACTERIAL LEUCYL AMINOPEPTIDASE PRECURSOR (EC 3.4.11.10).
OS Vibrio proteolyticus (Aeromonas proteolytica).
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 107-136 AND 233-405.
RC STRAIN=AFCC 15338;
RX MEDLINE: 92329552.
RA VAN HEKE G., DENSLON S., WATKINS J., WILSON K., WAGNER F.;
RT "Cloning and nucleotide sequence of the vibrio proteolyticus
RT aminopeptidase gene";
RL Biochim. Biophys. Acta 1131:337-340(1992).
RN [2]
RP SEQUENCE OF 73-504 FROM N.A.
RX MEDLINE: 92235063.
RA GUENET C., LEPAGE P., HARRIS B.A.;
RT "Isolation of the leucine aminopeptidase gene from Aeromonas
RT proteolytica. Evidence for an enzyme precursor";
RL J. Biol. Chem. 267:8390-8395(1992).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 107-396.
RX MEDLINE: 94373500.
RA CHEVRIER B., SCHALK C., D'ORCHYMONT H., RONDEAU J.M., MORAS D.,
RA TARNOUS C.;
RT "Crystal structure of Aeromonas proteolytica aminopeptidase: a
RT prototypical member of the co-catalytic zinc enzyme family.";
RL Structure 2:283-290(1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY COMPLEXED WITH INHIBITOR.
RX MEDLINE: 96213434.
RA CHEVRIER B., D'ORCHYMONT H., SCHALK C., TARNOUS C., MORAS D.;
RT "The structure of the Aeromonas proteolytica aminopeptidase complexed
RT with a hydroxamate inhibitor. Involvement in catalysis of Glu151 and
RT two zinc ions of the co-catalytic unit";
RL Eur. J. Biochem. 237:393-398(1996).
CC -1- CATALYTIC ACTIVITY: ACTS MOST RAPIDLY ON L-LEUCYL-PEPTIDES.
CC AMIDE AND BETA-NAPHTHYLAMIDE. DOES NOT CLEAVE GLU- AND ASP- BONDS.
CC -1- COFACTOR: BINDS TWO ZINC IONS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M28.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z11993; CAAT8039.1; -
DR EMBL: M85159; AAA21940.1; -
DR PIR: S24314; S24314.
DR PIR: S21684; S21684.
DR PDB: 1AMP; 31-AUG-94.
DR PDB: 1IG8; 01-AUG-96.
KM Hydrolyase; Aminopeptidase; Signal; Zymogen; Zinc; 3D-structure.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 106 POTENTIAL.
FT CHAIN 107 405 BACTERIAL LEUCYL AMINOPEPTIDASE.
FT PROPEP 406 504 REMOVED IN MATURE FORM (POTENTIAL).
FT METAL 211 211 ZINC (POTENTIAL).
FT METAL 214 214 ZINC (POTENTIAL).
FT METAL 336 336 ZINC (POTENTIAL).
FT DISULFID 329 333

FT ACT SITE 257 257 GENERAL BASE
FT CONFLICT 303 304 TD -> DT (IN AA SEQUENCE).
FT CONFLICT 306 306 N -> D (IN AA SEQUENCE).
FT CONFLICT 312 314 TOL -> QT (IN AA SEQUENCE).
SQ SEQUENCE 504 AA; 54232 MW; 36911698 CRC32;

Query Match 3.4%; Score 143; DB 1: Length 504;
Best Local Similarity 23.1%; Pred. No. 0.0063;
Matches 72; Conservative 37; Mismatches 105; Indels 98; Gaps 16;

QY 344 PVY-----ASSGSPAPQISADIAS-----RLKLKLGPAV-----PEMQ-----GSL 385
DB 108 PPTIQQATVTAMLPQVDASGITISLSEFTNFTYTTGGAQSDIASWQALSL 167
QY 386 GSPYHLCGPRRLVYNNHSTPNNIFCIEGRSEPDHYVYGAORDAW----- 437
DB 167 -----PNASVQVSHSGYNQ---KSVMTITGSEAPDEMIYIGHLDTIGSHTEQS 216
QY 437 -GPGAASAVGTAILLELVTFSSWNSNGFRPRSLFLISWDGDDGSGVSTEMLECYLS 495
DB 217 VAPGADDADAGIAVTEVIVLS---ENNFPKRSIAFMAYAAEEVGLRGSODLANQY-- 272
QY 496 VLHKAAVYVSLDNAVLDGDKFPAKTSPLTSLSEVLRQVDSPNHSGOTLREQVFTNP 555
DB 272 -----KSECKN-----VVSAL-GLDMTKRGS- -DDVFEITD 301
QY 556 SMDAEVIRPLP-----MDSAYSSTAF-----VGPAV- -FFSEMEDDOAY 594
DB 302 YTDNFQOYLQMLDELPLSLITGFDICGYACSDHASMNAAGYPAAMPFSKFNQY---- 359
QY 595 PFLTKEDTYEN 606
DB 359 PRHTTODTLAN 370

RESULT 6
SUBV_BACSU
ID SOBV_BACSU STANDARD; PRT; 806 AA.
AC P29141;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MINOR EXTRACELLULAR PROTEASE VPR PRECURSOR (EC 3.4.21.-).
GN VPR OR IPA-45R.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 161-195.
RX MEDLINE: 92041574.
RA SLOMA A., RUFO G.A. JR., THERIAULT K.A., DWYER M., WILSON S.W.,
RA PERO J.;
RT "Cloning and characterization of the gene for an additional
RT extracellular serine protease of Bacillus subtilis";
RL J. Bacteriol. 173:6889-6895(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE: 95020537.
RA GLASER P., KUNST F., ARNAUD M., COUDART M.P., GONZALES W.,
RA HULLO M.F., IONESCU M., LOBOCHINSKY B., MARCELINO L., MOSZER I.,
RA PRESECAN E., SANTANA M., SCHNEIDER E., SCHWEIZER J., VERTES A.,
RA RAPPOPORT G., DANCIN A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees";
RL Mol. Microbiol. 10:371-384(1993).
CC -1- FUNCTION: NOT REQUIRED FOR GROWTH OR SPOULATION.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- PTM: PROBABLY UNDERGOES C-TERMINAL PROCESSING OR PROTEOLYSIS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----

FT DOMAIN 197 314 COIL 2A.
 SQ SEQUENCE 1805 AA; 198744 MW; 789C2B48 CRC32:

Query Match 2.7%; Score 112.5; DB 1; Length 1805;
 Best Local Similarity 22.0%; Pred. No. 7.2;
 Matches 106; Conservative 44; Mismatches 149; Indels 183; Gaps 27;

QY 16 PSSQTVYRVGPRK---GHLEEEDEDEG-----AETLAHFCMEIRGPEPLGSR 67
 DB 1447 PAGESALEERVENEPFEGEIPGQDWEEGRESEADLGETLP---DSTPLDLYR 1502
 QY 68 QNLLPFWAAG--RRAPYLVTALTIFGAFLLGYAFRGSCQAGDS-----VLVYS 119
 DB 1503 SPASPMWDLAGEORLSPQ-----GDAKEDMGPAVPPA 1535
 QY 120 EDVNEPDLDFHQRLYMSDLQAMFLQFLGRLDPTIRQTSIRERVAGSAGMALUTDI 179
 DB 1536 QGLSGPEEEEG--HGSDLSS-----EEFDLGEASLLPGV----- 1573
 QY 180 RAALSQKLDHWTDTNHYGLQEPDPAHNTLHW-----VDEAKYGEQLPLED 228
 DB 1573 ---PEVADHVG-----OVPPYLQAPC--WDQGESDGFADDEEGEGEE---BD 1615
 QY 229 PUYCPYSAIGNVTGELYAH--YGRPEDLDLRARGVDPVGRLLVRGVISFAQKVN 286
 DB 1616 AD-----EGASGQMGWMSGASGGCKVQDIAQKG-DPYVE---SVG-----VSG 1657
 QY 287 ADFGAQGVLT-YPEPADFSQDPKPKSL-----SQAAVYGHV--HL-----G 326
 DB 1658 LWDGDLRGAAAVPALEWVSQDSAPSGSESEASLSEEGQVTDHLDAPQEVTSWVG 1717
 QY 327 TGDPTPG--PFSFNQTPPPVASSGLSPISPAQPI-SADIASLLRK-----LKGPVP 377
 DB 1718 VDAEDIGQSPNLSDEOVNKGKMGLEQAEQVLDGDELDLQGEVGAALKVP--- 1775
 QY 378 QEMQSLGSPYHLGPGPRLVNVNHRSTPINNIFGCIEGSEPDHVVIGAQDANG 437
 DB 1775 -----LVASPVHLGP-----SPLKTTLSGVDSMS 1801
 QY 438 PG 439
 DB 1802 SG 1803

RESULT 10
 P2P_LACPA STANDARD: PRT; 1902 AA.

AC Q02470;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-
 ASSOCIATED SERINE PROTEINASE) (Lp151).
 GN PRTP.
 OS Lactobacillus paracasei.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC Lactobacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCDO 151;
 RX MEDLINE; 92381481.
 RA HOLCK A., NAES H.;
 RA "Cloning, sequencing and expression of the gene encoding the cell-
 envelope-associated proteinase from Lactobacillus paracasei subsp.
 RT paracasei NCDO 151".
 RL J. Gen. Microbiol. 138:1353-1364(1992).
 RN [2]
 RP SEQUENCE OF 189-196.
 RX MEDLINE; 92226694.
 RA NAES H., NISSEN-MEYER J.;
 RA "Purification and N-terminal amino acid sequence determination of the
 cell-wall-bound proteinase from Lactobacillus paracasei subsp.

FT paracasei".
 RL J. Gen. Microbiol. 138:313-318(1992).
 CC -1- FUNCTION: PROTEASE WHICH BREAK DOWN MILK PROTEINS DURING THE
 CC GROWTH OF THE BACTERIA ON MILK.
 CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
 CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
 CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
 CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
 CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
 CC INSULIN B-CHAIN.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; M83946; AAA25248.1; -.
 CC PIR; B44858; B44858.
 CC HSSP; 099405; IMPT.
 CC PRAM; PF00746; Gram_pos_anchor; 1.
 CC PRAM; PF00746; Gram_pos_anchor; 1.
 CC PROSITE; PS00136; SUBTILASE_ASP; 1.
 CC PROSITE; PS00137; SUBTILASE_HIS; 1.
 CC PROSITE; PS00138; SUBTILASE_SER; 1.
 CC PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE NEG.
 CC KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal;
 CC Transmembrane; 1.
 CC FT SIGNAL 1 33
 CC FT PROPEP 34 187
 CC FT CHAIN 188 1902
 CC FT DOMAIN 188 1876
 CC FT TRANSMEM 1877 1895
 CC FT DOMAIN 1896 1902
 CC FT ACT_SITE 217 217
 CC FT ACT_SITE 281 281
 CC FT ACT_SITE 620 620
 CC FT DOMAIN 1867 1872
 CC FT PROTEINS
 CC FT SEQUENCE 1902 AA; 200253 MW; 932EC634 CRC32;

Query Match 2.7%; Score 111.5; DB 1; Length 1902;
 Best Local Similarity 21.3%; Pred. No. 9.2;
 Matches 100; Conservative 63; Mismatches 164; Indels 143; Gaps 23;

QY 214 VDEAKVGEQL-----PLEDDVYCPYSAIGNVTGELYA----- 249
 DB 335 IEDSAKIGADVLLNMSLGSQNGTLEDEPIAAVQANESGTAAVASGNSGTSGATGV 394
 QY 249 ---HYGREDDQLDRLARGVDPVGRLLVRGVISPAQKVTNAQ--DFAQGVLYLPEPAD 303
 DB 395 NKDYTG---LQDNEMVGT-----GTSRCATTVASAENTDVLSQAVTI--TDKRD 439
 QY 304 FSDPPKPSLSQQAIVYGHVHLGTDPYTPGFSFNQTPPPV--ASSGLSPISPAQPIA 361
 DB 440 LQLGPEITQLSNDR-----TG-----SPDQKKFFVVDASGDSLKGAAADYTA 483
 QY 362 DIASRLKLKLGYPVAPQEMQSLGSPYHLGPGPRLVNVNHRSTPINNI----- 414
 DB 484 DAKGIAIVKRGELNFAQKQ-----KYAQAAAGAGLIIVNNDGATGATLSIRLTTPPT 537
 QY 414 FGCIGRSEPDHVVYIGQRAMGGAASAVGATILILEVRT-----ESSMYSNFR 466
 DB 538 FG-ISSKT-----GQKLVDMWTAHPDSDLGVKIALTLPLNOKYTEDKMSDFSTYG-- 587
 QY 467 PRRLSLF---LSWDGDDGVSQSTEWLEGYLS-----VLHKAIVVYVSDNNAVIG 513

```

DB 587 PVSNTLSEKPDITAPG---GNINSTONNGYTNNGSTMSAPFIAGSOALLKQALNK---- 641
OY 514 DDKFAKTSPLLSLIESVLKQVD-----SNHGQGLIEVVTNPS 556
DB 641 NMFEDYQOLKGTALTDLEKTVEMTAQPIINDINNNVIVSRGAGI---VDVKA 696
OY 557 WDAEVRIRPLMDSSAFYAFVPAVEF-SFEMDDOAVFLTKDTE 605
DB 697 IDA-----LEKNPSTVVAENGYPVELKDFSTDKTRFLITNTTTH 739

RESULT 11
SCPA_STRPY
ID SCPA_STRPY STANDARD: PRT: 1167 AA.
AC P15926;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE C5A PEPTIDASE PRECURSOR (EC 3.4.21.-) (SCP).
GN SCPA.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-41.
RX MEDLINE: 90153964.
RA CHEN C.C., CLEARY P.P.;
RT "Complete nucleotide sequence of the streptococcal C5a peptidase gene
RT of Streptococcus pyogenes."
RL J. Biol. Chem. 265:3161-3167(1990).
CC -1- FUNCTION: THIS VIRULENCE FACTOR OF S. PYOGENES SPECIFICALLY CLEAVES
CC THE HUMAN SERUM CHEMOKIN C5A AT LYS(68)-ASP(69) BOND NEAR ITS
CC C-TERMINUS, DESTROYING ITS ABILITY TO SERVE AS A CHEMOTRACTANT.
CC -1- SUBCELLULAR LOCATION: CELL WALL BOUND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: J05229; AAA26960.1;
DR PIR: A35066; A35066.
DR HSSP: P00782; 2SBR.
DR PFAM: PF00082; Peptidase_S8.
DR PFAM: PF00746; Gram_pos_anchor.1.
DR PROSITE: PS00136; SUBTILASE_ASP.1.
DR PROSITE: PS00137; SUBTILASE_HIS.1.
DR PROSITE: PS00138; SUBTILASE_SER.1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING: FALSE_NEG.
KW Hydrolase; Serine protease; signal; Cell wall; Repeat; Transmembrane.
FT SIGNAL 1 31
FT CHAIN 32 1167 C5A PEPTIDASE.
FT DOMAIN 32 1139 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1140 1157 POTENTIAL.
FT DOMAIN 1158 1167 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 130 130 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 193 193 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 512 512 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1029 1104 CELL WALL ATTACHMENT DOMAIN (POTENTIAL).
FT DOMAIN 1034 1101 4 X 17 AA TANDEM REPEATS.
FT REPEAT 1051 1050 1.
FT REPEAT 1051 1067 2.
FT REPEAT 1068 1084 3.
FT REPEAT 1085 1101 4.
FT DOMAIN 1127 1131 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1167 AA; 128263 MW; 58166B7B CRC32;

```

```

Query Match 2 6%; Score 108.5; DB 1; Length 1167;
Best Local Similarity 18.8%; Pred. No. 7.3;
Matches 139; Conservative 90; Mismatches 227; Indels 285; Gaps 37;

OY 30 RKGHLEEEDEDEGEAGETLAHFCPMLRGPEPL-GSRPQPNLIPMAAGRAAPLVLT 88
DB 183 KQKTAIVDQEHGTHSGILSGNAPSETKEPYRLEGAMPE-----AQLLLM 227
OY 89 ALLIFGAFILGYVAFRGSCQAGDSVLVSDVNYEPDLDFHOGRLYWSDLQAMFLQFL 148
DB 228 RVEIYNG--LADYA--RNAQAIRDAVNLGARVIN---MSGNALAVANLP----- 273
OY 149 GEGRELDITRQSLRERVAGSGMAALITODITAAISROKLDHWIDTHYVG-LQFPDPAH 207
DB 273 -----DETRKAFDYAKSKGVSIVTS-----AGNDSFGGKTRFLPAH 310
OY 208 PNTLHWVDAGYVG-----EDLPLEDPPVYCP 234
DB 311 P-----DYGVGIPPAADSTLTVASISPPKQLETANVKTDQDQKEMPVLSTRFEF 363
OY 235 YSAIGNVTGELYVAHYGRPEDLQDLRARGVDVGRLLVRVGVISFAQKVTNAODFGAOG 294
DB 364 -----NKAYDYAVANGMKED--DFK-----DYKGIALLERDIDIDKDKVANA KKGAVG 412
OY 295 VLIIPE-----PADF-----SQDPPKPSLSQQAAYGVHVLGTGDPY 331
DB 413 VLIYDQDKGFPIELPNVDQMAAFISRKDGLLLKDNPKQITTFNA----- 459
OY 332 TGC-FPSFNQOTFPVYASGL-----PSI--PAQPIADIASRLRLKGPVAPQEWQ 381
DB 459 TPRLVFTASGTRKLSRFSWGLTADGNIKPDIAAGODILSVANNKYALSGTSMAPLV 518
OY 382 GSLLG-----SPYHLGPG----- 395
DB 519 AGIMGLQKQYETQYPDMPSERLDLAKKVLMSATALLVDEDEKAFSPRQGAGAVDAK 578
OY 395 ---PRRLVYNNHRSTPL--NNIFG-----CIEGRSE--PDHYVITGAQRD----- 435
DB 579 KASATMYVTYDKDNMISSKVLNVSDFEVTYVHNKSDKPOLYYQATVQTDKVDGKHF 638
OY 435 AMQPGA-----AKSAGVTAIILEIVRTS---SNWSNPF-----RP 467
DB 639 ALAPKVLYPASWQKITIPANSSKQVTPIDASR-FSKDLIAQKNQYFLEGVYRFKQDPT 697
OY 468 RRSLLFISWDG--GDFGSYGSTEMLEGYSLVHLKAVYVSLDNAVIGDDKFAKTSPLL 525
DB 698 KEELMSIPYIGFPGDGFGNLSAVE-----KPIYDSKD-----GSSYHEANSDAK 741
OY 526 TSLIESVLKQVDSPNHSGOTLYEQVFTNPMSDAEYIRPL-----PMDSASFTAF 577
DB 742 DQLDGDGLQFYALKNNFTALTTE---SNP-W--TIKAVKEGVENIEDIESESEITETIF 794
OY 578 VGVPAVEFSFEMDDOAVPFLH 598
DB 795 AGTEFA-----KODDSHYITH 810

RESULT 12
EFG2_TREPA
ID EFG2_TREPA STANDARD: PRT: 695 AA.
AC 083748;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ELONGATION FACTOR G 2 (PF-G 2).
GN FUSB OR FUSA-2 OR TP0767.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NICHOLS.
RX MEDLINE: 98332770.

```

```

RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA DODSON R., GINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA SODERGEN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA HATCH B., HORST K., ROBERTS K., MATTHEY L., WEIDMAN J., SMITH H.O.,
RA VENTER J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT Spirochete."
RL Science 281:375-388(1998).
CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC RIBOSOME.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001248; AAC65735.1;
DR TIGR; TP0067;
DR PFAM; PF00009; GTP_EFTU; 1.
DR PFAM; PF00679; EFG_C; 1.
DR PROSITE; PS00301; EFATOR.GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding;
KW Multitene family;
ET NP_BIND 15 22 GTP (BY SIMILARITY);
ET NP_BIND 82 86 GTP (BY SIMILARITY);
ET NP_BIND 136 139 GTP (BY SIMILARITY);
FT FT
FT SEQUENCE 695 AA; 76831 MW; 58697234 CRC32;
SO
Query Match 2.6%; Score 107.5; DB 1; Length 695;
Best Local Similarity 21.1%; Pred. No. 3.9; Mismatches 167; Indels 169; Gaps 24;
Matches 105; Conservative 56;

```

```

OY 484 VGSTEWE---GYLSVL 497
Db 638 VDAEVPLEMEGFSTVL 654
RESULT 13
ID UFO_HUMAN STANDARD; PRT; 887 AA.
AC P30530;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DE TYROSINE-PROTEIN KINASE RECEPTOR UFO PRECURSOR (EC 2.7.1.112) (AXL
DE ONCOGENE).
GN AXL OR UFO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RX MEDLINE; 92050809.
RA JANSSEN J.W.G., SCHULZ A.S., STEENVOORDEN A.C.M., SCHMIDBERGER M.,
RA STREHL S., AMBROS P., BARTRAM C.R.;
RT "A novel putative tyrosine kinase receptor with oncogenic potential.";
RL Oncogene 6:2113-2120(1991).
RN [2]
RX MEDLINE; 92017777.
RA O'BRYAN J.P., FRYE R.A., COGSWELL P.C., NEUBAUER A., KITCH B.,
RA PROKOP C., ESPINOSA R., LE BEAU M.M., EARP H., LIU E.T.;
RT "AXL, a transforming gene isolated from primary human myeloid
RT leukemia cells, encodes a novel receptor tyrosine kinase.";
RL Mol. Cell. Biol. 11:5016-5031(1991).
RN [3]
RX MEDLINE; 94067791.
RA LEE S.T., STROUK K.M., SPRITZ R.A.;
RT "A survey of protein tyrosine kinase mRNAs expressed in normal human
RT melanocytes.";
RL Oncogene 8:3403-3410(1993).
CC -1- FUNCTION: MAY FUNCTION AS A SIGNAL TRANSDUCER BETWEEN SPECIFIC
CC CELL TYPES OF MESODERMAL ORIGIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DISEASE: HAS TRANSFORMING POTENTIAL IN PATIENTS WITH CHRONIC
CC MYELOPROLIFERATIVE DISORDER OR CHRONIC MYELOIDYTIC LEUKEMIA.
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57019; CAA40338.1; ALF_INIT.
DR EMBL; M76125; AAA61243.1; ALF_INIT.
DR EMBL; S65125; AAB20305.1; ALF_INIT.
DR HSSP; P11362; 1FGI.
DR MIM; 109135;
DR PFAM; PF00041; fn3; 2.
DR PFAM; PF00047; 1g; 2.
DR PFAM; PF00069; pkinase; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;
KW Transferase; Phosphorylation; Transmembrane; Signal; Repeat;
KW Immunoglobulin domain; Proto-oncogene; Alternative splicing.

```

FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 887 TYROSINE-PROTEIN KINASE RECEPTOR UFO.
FT DOMAIN 19 442 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 443 465 POTENTIAL.
FT DOMAIN 466 887 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 29 118 IG-LIKE DOMAIN.
FT DOMAIN 134 205 IG-LIKE DOMAIN.
FT DOMAIN 217 315 FIBRONECTIN TYPE-III.
FT DOMAIN 319 416 FIBRONECTIN TYPE-III.
FT DOMAIN 529 800 PROTEIN KINASE.
FT NP_BIND 535 543 ATP (BY SIMILARITY).
FT BINDING 560 560 ATP (BY SIMILARITY).
FT ACT_SITE 665 665 BY SIMILARITY.
FT MOD_RES 696 696 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT VARSPLIC 422 430 MISSING (IN SHORT FORM).
FT CONFLICT 296 296 P -> T (IN REF. 2).
FT CONFLICT 331 331 E -> K (IN REF. 2).
FT CONFLICT 632 632 G -> D (IN REF. 2).
SQ SEQUENCE 887 AA; 97374 MW; E747BA73 CRC32;

Query Match 2.5%; Score 107; DB 1; Length 887;
Best Local Similarity 21.1%; Pred. No. 6.2;
Matches 118; Conservative 66; Mismatches 142; Indels 232; Gaps 33;

QY 282 QKVTNAQDFGAGVLIYPPADFSODPPKPSLSQQAAYGH-VHGTGDPYTPGF----- 336
DB 254 QAVLSDDGNGIOA---GSP-----DPPEPLTSQASVPPHOLRLGSLPHPPYHIRVAC 304
QY 336 -----PSFNQTOFPYVAS-SGLPSIAPDISAD-INSRLRLKLGVARQEMOGSLGSP 388
DB 305 TSSQGPS-SWTWLPVETVEGVPLGPENISATRNGSQAFVHQEPRAP--LQGITLG-- 360
QY 389 YHGG---GPRRLRVVNNHRTST-----PINNIFGCIENRSEPDHYVIG----- 431
DB 360 YRLAAGQOTPRVMDIGLRQEVTTLELQDGSVSNLTVCVAA-----YTAGDGPWSLP 413
QY 431 AORDAMGPAAK-----SANGTALILELV----- 455
DB 414 VPLEAMRPEADPVHQLVKEPSTPAFSWPMWYVLLGAAVVAACVLITALELVHRRKKEFR 473
QY 455 -RTFSWMSNG-----FPRRSRL-----LFISM----- 478
DB 474 YGEVEPTVERGELVYRVKRSYSKRTTEATINSIGISELKEKRDVMDRHKYALK 533
QY 478 ---GGDFGSGVSTEMLEGYLVHLKAVYVYSLDNNALVLDGDKFHART-----SPLLT 526
DB 534 TLGEGFGAV-----MEGL-----NODDSIL--KVAVTKMIAICTRELEED 574
QY 527 SLIESV-LKQVSPNHSQGLYEQVFTNPMDAIEVRLPM---DSAYSTFAEVGVA 582
DB 575 FLSEAVCMKEFDHPN-----VMRLIGVCFQOSESESESPA---FV 610
QY 583 VEESFEMEDQAVPELHTEKEDYENLHKVLOGRLPAVAQAAGOLLRLSHDRLLPD 642
DB 611 VILPFAKHDDLISFL-----LISRLGGQ-----PVILPTMLVKNADIASGWE 654
QY 643 FGR---YGDVYLRHIGNLNE-----FSGDLKARG---LTLQWYSAR 678
DB 655 YLSTRKFIHRLAARNC-MLNEMMSVCVADFGLSKRIYNGDYRQRIKMPKWI----- 710
QY 679 GDYIRAEKLRQETISSE 696
DB 710 -----AIESIADRVTYSK 722

RESULT 14
PIP_LACLC STANDARD: PRT: 1902 AA.
AC P16271;
DT 01-APR-1990 (rel. 14, Created)
DT 01-APR-1990 (rel. 14, Last sequence update)
DT 01-NOV-1995 (rel. 32, Last annotation update)

DE PI-TYPE PROTEINASE PRECURSOR (EC 3.4.21.-) (WALL-ASSOCIATED SERINE
DE PROTEINASE).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pMV05.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-WG2;
RC MEDLINE: 88149035.
RA KOK J., LEENHOUTS K.J., HAANDRIKMAN A.J., LEDEBOER A.M., VENEMA G.;
"Nucleotide sequence of the cell wall proteinase gene of
Streptococcus cremoris Wg2".
RL Appl. Environ. Microbiol. 54:231-238(1988).
CC -1- FUNCTION: PROTEASE WHICH BREAK DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC INSULIN B-CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M24767; AAA17677.1; -.
DR HSRP: Q99405; IMPT.
DR PFMAM: PF00082; Peptidase_S8; 3.
DR PFMAM: PF00746; Gram_pos_anchor; 1.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KM Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KM Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1902 AA; 199910 MW; FDE60D93 CRC32;

Query Match 2.5%; Score 106.5; DB 1; Length 1902;
Best Local Similarity 20.3%; Pred. No. 21;
Matches 96; Conservative 65; Mismatches 163; Indels 149; Gaps 22;

QY 214 VDEAGKVGSQL-----PLEDDPYVCYPAIGNVTGELYVA----- 249
DB 335 IEDSAKIGADVLMKSGSDSGNOTLEDPFLAIVQANESGTAIVASGSGTSGETGV 394
QY 249 ---HGKRPEDLDDLRARGVDPVGR-----LLIVRGVYSFAQVTNADFGAGVLIYPE 300
DB 395 NKDYDG---LQDNEVWGPGTSRGATVVASAENTDVITQAVTIND---GTGLQLGP- 445
QY 301 PADFSODPPKPSLSQQAAYGHVHGTGDPYTPGFPSFNQTOFPYV--ASSLPSIPAP 358

```

Db 445 -----GTQLSSND-FT---GSFDQKFFVVKDASGNLSKALAD 480
QY 359 ISADIASRLRLKLGPAVPEWQSLGSPYHLGPGRLVYNNHRTSPINNI-----414
Db 481 YTAADKAGKIAIVKRGELTFADKQ-----KYAQAAGAGLIIVNNDGTATPVTSMALLTT 534
QY 414 ---FGCIEGRSEPDHYVYVIGAORDAMGPAKSAVGAIIILEVRT-----FSSVSN 463
Db 535 FPTFGLS-----VTGOKLYDWTVAHDDSLGVKIALTLVPNOKYTEDKMSDFTSY 585
QY 464 GFRPRRSLF---ISWDGDFGSGVSTEMLEGYS-----VLHKAVVYVSLDNA 510
Db 586 G--PVSNLSEFKPDITAG---GNIMSTQNNNGYTNMGSTMASPFIAQSALLKQALNKK 640
QY 511 VLGDGDKFHAKTSPLLTSLIESVLKQVD-----SPNHSQOTLYEQVYFT 553
Db 641 ---NNPFYAYYKOLKGTALDFTLKVEMNTAOPINDINYNVIVSPRQAGL-----VDV 693
QY 554 NPSMDAEVIRPLMDSSAYSFATFVGPVAFEF-SFMEDDQAYFHLTKEDTYE 605
Db 694 KAALDA-----LEKNPSTVAENGIPAVELKDFSTDKFTKLTFTNRTTHE 739

RESULT 15
P2P_LACLA STANDARD: PRT: 1902 AA.
ID P2P_LACLA
AC p15293:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-
DE ASSOCIATED SERINE PROTEINASE) (LP151).
GN PRT.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OG Plasmid pLP763.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCDO 763.
RX MEDLINE; 89313288.
RA KIMWAKI M., IKEMURA H., SHIMIZU-KADOTA M., HIRASHIMA A.;
RT "Molecular characterization of a cell wall-associated proteinase gene
RT from Streptococcus lactis NCDO763."
RL Mol. Microbiol. 3:359-369(1989).
CC -I- FUNCTION: PROTEASE WHICH BREAK DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -I- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC SPECIFICITY, ALTHOUGH SOME SUBSTRATE PREFERENCE HAVE BEEN NOTED,
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC INSULIN B-CHAIN.
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X14130; CAA32350.1; -.
CC PIR; S06997; S06997.
CC HSSP; P00782; 2SBT.
CC FRAM; P00082; Peptidase_S8; 3.
CC DR FRAM; P00746; Gram_pos_anchor; 1.
CC DR PROSITE; PS00136; SUBTILASE_ASP; 1.
CC DR PROSITE; PS00137; SUBTILASE_HIS; 1.
CC DR PROSITE; PS00138; SUBTILASE_SER; 1.

```

```

DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1875
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
FT SEQUENCE 1902 AA; 200139 MW; 8617ECFB CRC32;

Query Match 2.5%; Score 106.5; DB 1; Length 1902;
Best Local Similarity 20.3%; Pred. No. 21;
Matches 96; Conservative 62; Mismatches 166; Indels 149; Gaps 21;

QY 214 VDEAGKVEQL-----PLEDDPYCYSAIGNYTGELYVA-----249
Db 335 IEDSAKIGADVLMNLSGSDGNQTLDEPDLAAVQANESGTAIVISAGNGSGSATEGV 394
QY 249 ---HYGRPEDLODLRANGVPVGR-----LLIYRVGYISTAQVYTNADFGAAGVLIPE 300
Db 395 NKDIYG---LQDNEMVGTPTSGRGTVAESAENTDITVAVTITD---GTGLQIGPE 445
QY 301 PADFSODPPKPSLSSQOAVYGHVHLGTDPYTPGPFNFOTOPPPV--ASSGLPSIPAOP 358
Db 446 TIQLSSND-----FTG-----SFDQKFFVVKDASGNLSKGVAD 480
QY 359 ISADIASRLRLKLGPAVPEWQSLGSPYHLGPGRLVYNNHRTSPINNI-----414
Db 481 YTAADKAGKIAIVKRGELTFADKQ-----KYAQAAGAGLIIVNNDGTATPVTSMALLTT 534
QY 414 ---FGCIEGRSEPDHYVYVIGAORDAMGPAKSAVGAIIILEVRT-----FSSVSN 463
Db 535 FPTFGLS-----VTGOKLYDWTVAHDDSLGVKIALTLVPNOKYTEDKMSDFTSY 585
QY 464 GFRPRRSLF---ISWDGDFGSGVSTEMLEGYS-----VLHKAVVYVSLDNA 510
Db 586 G--PVSNLSEFKPDITAG---GNIMSTQNNNGYTNMGSTMASPFIAQSALLKQALNKK 640
QY 511 VLGDGDKFHAKTSPLLTSLIESVLKQVD-----SPNHSQOTLYEQVYFT 553
Db 641 ---NNPFYAYYKOLKGTALDFTLKVEMNTAOPINDINYNVIVSPRQAGL-----VDV 693
QY 554 NPSMDAEVIRPLMDSSAYSFATFVGPVAFEF-SFMEDDQAYFHLTKEDTYE 605
Db 694 KAALDA-----LEKNPSTVAENGIPAVELKDFSTDKFTKLTFTNRTTHE 739

```

Search completed: January 13, 2000, 01:09:18
Job time: 135 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2000, 12:10:22 ; Search time 39.36 Seconds

(without alignments)
1246.581 Million cell updates/sec

Title: US-09-358-755-1

Sequence: 1 MERLWGLFORAQLSPRSSQ.....TLGGANALSGDWMINDNF 801

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database: SPTREMBL_11:*

Word size: 0

Number of hits that pass the threshold: 199794

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|--------|--------------------|
| 1 | 4034.5 | 96.0 | 780 | 4 | 075422 | 075422 homo sapien |
| 2 | 1504.5 | 35.8 | 757 | 11 | 007891 | 007891 cricetus |
| 3 | 1450.5 | 34.5 | 763 | 11 | 062351 | 062351 mus musculu |
| 4 | 1424.5 | 33.9 | 622 | 11 | 099376 | 099376 rattus norv |
| 5 | 1349.5 | 32.1 | 778 | 13 | 090997 | 090997 gallus gall |
| 6 | 775 | 18.4 | 752 | 11 | P70627 | P70627 rattus norv |
| 7 | 756 | 18.0 | 745 | 11 | 054697 | 054697 rattus norv |
| 8 | 727 | 17.3 | 750 | 4 | 043748 | 043748 homo sapien |
| 9 | 722 | 17.2 | 752 | 11 | 035409 | 035409 mus musculu |
| 10 | 709.5 | 16.9 | 751 | 6 | 077564 | 077564 sus scrofa |
| 11 | 621 | 14.8 | 751 | 5 | P91406 | P91406 caenorhabd |
| 12 | 541 | 12.9 | 794 | 3 | 043023 | 043023 schizosacch |
| 13 | 510 | 12.1 | 1483 | 5 | 093332 | 093332 caenorhabd |
| 14 | 297.5 | 7.1 | 783 | 3 | 008919 | 008919 saccharomyc |
| 15 | 271.5 | 6.5 | 277 | 11 | 061560 | 061560 mus musculu |
| 16 | 243 | 5.8 | 809 | 3 | 008693 | 008693 saccharomyc |
| 17 | 234.5 | 5.6 | 322 | 4 | 043176 | 043176 homo sapien |
| 18 | 208 | 4.9 | 63 | 11 | 064187 | 064187 cavia porce |
| 19 | 159 | 3.8 | 501 | 2 | P96152 | P96152 vitrio chon |
| 20 | 155.5 | 3.7 | 493 | 5 | 076552 | 076552 acanthochei |
| 21 | 154.5 | 3.7 | 485 | 2 | 053737 | 053737 streptomyce |
| 22 | 134.5 | 3.2 | 472 | 11 | 0921Y1 | 0921Y1 rattus norv |
| 23 | 132 | 3.1 | 393 | 2 | 082996 | 082996 aeromonas c |
| 24 | 131.5 | 3.1 | 565 | 1 | 057813 | 057813 pyrococcus |
| 25 | 128.5 | 3.1 | 835 | 2 | 069873 | 069873 streptomyce |

| | | | | | | |
|----|-------|-----|------|----|--------|---------------------|
| 26 | 126 | 3.0 | 566 | 5 | 093318 | 093318 caenorhabd |
| 27 | 120.5 | 2.9 | 500 | 2 | P96264 | P96264 mycobacteri |
| 28 | 117.5 | 2.8 | 374 | 3 | 004033 | 004033 saccharomyc |
| 29 | 117.5 | 2.8 | 561 | 3 | 074125 | 074125 coprinus ci |
| 30 | 116.5 | 2.8 | 1301 | 2 | 006307 | 006307 allicyclobac |
| 31 | 116.5 | 2.8 | 1083 | 2 | 086537 | 086537 streptomyce |
| 32 | 116 | 2.8 | 1017 | 2 | 005586 | 005586 mycobacteri |
| 33 | 115.5 | 2.7 | 640 | 3 | 002495 | 002495 schizophyll |
| 34 | 113.5 | 2.7 | 634 | 4 | 075114 | 075114 homo sapien |
| 35 | 112.5 | 2.7 | 701 | 12 | 083131 | 083131 avian myelo |
| 36 | 112.5 | 2.7 | 701 | 12 | 089810 | 089810 avian myelo |
| 37 | 112 | 2.7 | 1128 | 1 | 051999 | 051999 halobacteri |
| 38 | 112 | 2.7 | 433 | 11 | 070216 | 070216 mus musculu |
| 39 | 109.5 | 2.6 | 2731 | 2 | 0923T9 | 0923T9 pseudomonas |
| 40 | 108.5 | 2.6 | 1557 | 5 | 096652 | 096652 drosophila |
| 41 | 108.5 | 2.6 | 1075 | 11 | P97305 | P97305 mus musculu |
| 42 | 108 | 2.6 | 1150 | 2 | 053637 | 053637 streptococ |
| 43 | 107.5 | 2.5 | 809 | 5 | 044391 | 044391 strongyloce |
| 44 | 107 | 2.5 | 679 | 10 | 023352 | 023352 arabidopsis |
| 45 | 106.5 | 2.5 | 2314 | 2 | 069822 | 069822 streptomyce |

ALIGNMENTS

RESULT 1
ID 075422 PRELIMINARY; PRT; 780 AA.
AC 075422;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE TRANSFERRIN-RECEPTOR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RA GLOECKNER G., ROSENTHAL A., SCHERER S., WEBER J., SCHATTEVOY R.,
RA TSUI L.-C.;
RT "Large scale analysis of two regions in human chromosome 7q22:
RT annotation of 650 kb of genomic sequence around the PCOLCE and CUTL1
RT loci reveals 17 genes."
RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF053356; AAC78796.1; -
SQ SEQUENCE 780 AA; 86621 MW; 84BC2BD2 CRC32;

Query Match 96.0%; Score 4034.5; DB 4; Length 780.
Best Local Similarity 95.9%; Pred. No. 0;
Matches 774; Conservative 0; Mismatches 0; Indels 33; Gaps 2;
QY 1 MERLWGLFORAQLSPRSSQTVYQREVGPRKGLHEEEDGEGARTLAHFCPELRGPE 60
Db 1 MERLWGLFORAQLSPRSSQTVYQREVGPRKGLHEEEDGEGARTLAHFCPELRGPE 60
QY 61 PLGSRPQPLIIPMAAGRAAPYLVTALLITGAFLLGYVAFRSCQACGSVLVSE 120
Db 61 PLGSRPQPLIIPMAAGRAAPYLVTALLITGAFLLGYVAFRSCQACGSVLVSE 120
QY 121 DVNYEPDLDFHGRRLYSDIQAMFLOFLGEGRLIEDTIRQTSLEPRVAGSAGMALTRDIR 180
Db 121 DVNYEPDLDFHGRRLYSDIQAMFLOFLGEGRLIEDTIRQTSLEPRVAGSAGMALTRDIR 180
QY 181 AALSROKLDHVTDFTHVYGQFPDPAPNPNTLHVVDAGKVEQLPLEDDPVYPSAIGN 240
Db 181 AALSROKLDHVTDFTHVYGQFPDPAPNPNTLHVVDAGKVEQLPLEDDPVYPSAIGN 240
QY 241 VTGELYAAHGRPELDQDLFARGVDPVGRLLLRVGVISFAQKVTNAODFGAAGVLYPE 300
Db 241 VTGELYAAHGRPELDQDLFARGVDPVGRLLLRVGVISFAQKVTNAODFGAAGVLYPE 300
QY 301 PADFSODPKPRLSSQQAAYGVHLCTGDPYTPGPFPSFNOTOPPVASSGLPSIPAPDIS 360

```

Db 301 PADFSQDPKPSLSSQAVYGVHVLGTGDPYRGPFSFQNTQ----- 343
OY 361 ADIASRLRLKLVPAPOEMOGSLGSPYHLGPPRLVYNNHRTSPINNIFECIEGR 420
Db 343 -----KLKGPAPQEMOGSLGSPYHLGPPRLVYNNHRTSPINNIFECIEGR 393
OY 421 SEPDIYVIGAGRDAMGPAAGSAVGTALILELVYTFSSMNSNGRPRSLFLISWDGD 480
Db 394 SEPDIYVIGAGRDAMGPAAGSAVGTALILELVYTFSSMNSNGRPRSLFLISWDGD 453
OY 481 FGSVSTEMLEGLYSLVHLKAVYVSLDNAVYLGDKFNAKTSPLTSLIESYLKQVDSN 540
Db 454 FGSVSTEMLEGLYSLVHLKAVYVSLDNAVYLGDKFNAKTSPLTSLIESYLKQVDSN 513
OY 541 HSGQTLVEOVYFTNPSMDAEVIRPLPMDSSAYSTFAVGVPAVEFSFEMDDOAYPFLHTK 600
Db 514 HSGQTLVEOVYFTNPSMDAEVIRPLPMDSSAYSTFAVGVPAVEFSFEMDDOAYPFLHTK 573
OY 601 EDTYENLHKVLOGRLPAVAQAVAGLQGLLRSLSDRLPLDFGXYGDVYLRHIGNLNF 660
Db 574 EDTYENLHKVLOGRLPAVAQAVAGLQGLLRSLSDRLPLDFGXYGDVYLRHIGNLNF 633
OY 661 SSDLKARGITLQWYSANGDYIRAAEKLROEIIYSEERDELTIRMYNVRIMR-----VE 714
Db 634 SDDLKARGITLQWYSANGDYIRAAEKLROEIIYSEERDELTIRMYNVRIMRPLSAQVE 693
OY 715 FYFLSQYVSPADSPFRHIFPMGRGHTLGLALDLRLRLNSNGSGTGATISSTGFQESRFR 774
Db 694 FYFLSQYVSPADSPFRHIFPMGRGHTLGLALDLRLRLNSNGSGTGATISSTGFQESRFR 753
OY 775 QALTLTWLOGAANALSGDVWINDNPF 801
Db 754 QALTLTWLOGAANALSGDVWINDNPF 780

RESULT 2
007891 PRELIMINARY: PRT: 757 AA.
AC 007891:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE TRANSFERRIN RECEPTOR PROTEIN (TR).
OS Cricetus griseus (Chinese hamster).
OC Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetus.
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE-Ovary:
RX MEDLINE: 94012749.
RA COLLAM J.F., LAI A., DOMINGO D., FITCH M., HATTON S.,
RA TROMBRIDGE I.S.;
RT "YTRF is the conserved internalization signal of the transferrin
RT receptor, and a second YTRF signal at position 31-34 enhances
RT endocytosis";
RL J. Biol. Chem. 268:21686-21692(1993).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE: 90226333.
RA ALVAREZ E., GIRONES N., DAVIS R.J.;
RT "A point mutation in the cytoplasmic domain of the transferrin
RT receptor inhibits endocytosis.";
RL Biochem. J. 267:31-35(1990).
CC -1- FUNCTION: CELLULAR UPTAKE OF IRON OCCURS VIA RECEPTOR-MEDIATED
CC ENDOCYTOSIS OF LIGAND-OCCUPIED TRANSFERRIN RECEPTOR.
CC -1- SUBUNIT: DIMER OF SIMILAR OR IDENTICAL CHAINS LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- PTM: EACH CHAIN BINDS GLYCANS, PHOSPHATE, AND A PALMITATE GROUP.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
DR EMBL: L19142; AAA03576.1;
KW Transmembrane; Glycoprotein; Receptor; Lipoprotein; Palmitate;
Signal-anchor; Endocytosis.

```

```

FT DOMAIN 1 65 CYTOPLASMIC.
FT TRANSMEM 66 89 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (BY SIMILARITY).
FT FT DOMAIN 90 757 EXTRACELLULAR:
FT FT CONTAINS TRANSFERRIN BINDING SITE
FT FT (BY SIMILARITY).
FT FT SIZE 20 23 ENDOCYTOSIS SIGNAL.
FT FT DISULFID 89 89 INTERCHAIN (BY SIMILARITY).
FT FT DISULFID 97 97 INTERCHAIN (BY SIMILARITY).
FT FT CARBOHYD 248 248 BY SIMILARITY.
FT FT CARBOHYD 314 314 BY SIMILARITY.
FT FT CARBOHYD 719 719 POTENTIAL.
FT FT CARBOHYD 724 724 BY SIMILARITY.
FT FT CONFLICT 20 20 Y -> T (IN REF. 2).
FT SEQUENCE 757 AA; 85080 MW; A6A1749B CRC32;

Query Match 35.8%; Score 1504.5; DB 11; Length 757;
Best Local Similarity 42.8%; Pred. No. 3.9e-109;
Matches 334; Conservative 119; Mismatches 248; Indels 79; Gaps 19;

OY 35 EEEEDGEGAEETLAHFCPMELRGPELGRPPONLIPMAAGRAAPVLTALLIFT 94
Db 44 EENETDNMKASVGRK-----RLNGRLCEGT-----IAVIFFL 78
OY 95 GAFLIGYAF--RSCQAC-----CDLVYVSEOVNPEPDDFHFQGRLYNSDLQAMF 144
Db 79 IGFMIGYGYCKRKQKCVLAETGNSSEIIDEENP-----OSSRLYMDLKL 131
OY 145 LQFLGEGRLDITROTSLREHVASAGMALTDODIRALSRQKLDHWTDHYVGLQFPD 204
Db 132 SEKDIAIEFTTITQLOSTSRAGSQDENLAYIENQRFKLSKYWRDHYVQKYG 191
OY 205 PAHPNTHWVDEACKVGEOLPELDPDYVCYSAIGNVTGELVYAHYGRPEDLDLRAGV 264
Db 192 SAAQNAVITIIINVN--GSDLVENPGGVAYASKATVSGKLHANFGKKEFDLKY-- 246
OY 265 DPV--GRLLVAVGYISFAQKTNNAODGACGVLYLPEADISQPPKPSLSSQAVYGVH 323
Db 246 -PVNGSLIVRAGKITFEKAKNAQSFNAIGVLYMQTKF-----PVVAELSLFGHA 298
OY 324 HLGNGDPYTPGFSPFNQOPPPVASSGSLPAPDISADIRSLRLKLVPAPOEMOGS 383
Db 299 HLGNGDPYTPGFSPFNQOPPPVASSGSLPAPDISADIRSLRLKLVPAPOEMOGS 357
OY 384 LLGSPYHLGPPRLVYNNHRTSTPINNIFGCIEGRSEPDHYVIGAGRDAMGPAAGS 443
Db 358 SL-CKLESSQGINVNLVYNNVLKETRILNIFGVIKGEFEEDPRYIVVAGRDAMGPAAGS 416
OY 444 AVGTALILELVYTRSSMYS--NGFRPRSLFLISWDGDFGSGVSTEWLEGLYSLVHLKAV 502
Db 417 SVGTGLLKLQAQSDMYSRGGFPPRSRIIPASVAGDEGAVGATEWLEGLYSLVHLKAV 476
OY 503 VYVSLDNAVYLGDKFNAKTSPLTSLIESYLKQVDSNPHSQCOTLYEQVYFNPSMDAEVI 562
Db 477 TYINLDAKVLTRNFKVSAPLTYLTIETKQMDVRHP-IDCKPLXR-----DSMWISKV- 530
OY 563 RPLPMDSSAYSTFAVGVPAVEFSFEMDDOAYPFLHTKEDTYENL-HKVLOGRLPAVAQ 621
Db 530 EDLSLDNAVYLGDEYVSGIPAVSFWECEMED-YPLDTWLDYTERKLIQVPG--LNKMYRA 586
OY 622 VQAQAGLLRLSLSDRLPLDFGXYGDVYLRHIGNLNFSSDDLKARGITLQWYSARGDY 681
Db 587 AAEVAGQPIIKRLTHDIELNDYDNNKRLSFVELNFRADIRAMGSLQMLVSARDDF 646
OY 682 IRAEKLROEIIYSEERDELTIRMYNVRIMRVEYFLSQYVSPADSPFRHIFPMGRGHTL 741
Db 647 FRATSLTTDHNNAEKTRFYVREINNRIMKVEYHFLSPYSPESPRHATFMSGSHTL 706
OY 742 GALLDLRLRLNSNGSGTGATISSTGFQESRFRROLALTWTLOGAANALSGDVWINDNPF 801
Db 707 TALVENLKLRLKNS-----AFNETLRNQALATWTLOGAANALSGDVWINDNPF 757

```

```

RESULT 3
ID 062351 PRELIMINARY: PRT: 763 AA.
AC 062351;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)
DE TRANSFERRIN RECEPTOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BLXDBA/2; TISSUE=HEMATOPOIETIC;
RA THOMBRIDGE I.S.; DOMINGO D.L.; THOMAS M.L.; CHAIN A.;
RU Submitted (JAN-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; X57349; CAA40624.1;
SC SEQUENCE 763 AA; 85731 MW; 5C2B158C CRC32;

Query Match 34.5%; Score 1450.5; DB 11; Length 763;
Best Local Similarity 41.7%; Pred. No. 6.5e-105;
Matches 333; Conservative 127; Mismatches 257; Indels 81; Gaps 24;

QY 24 ORVGPGRKGHLE-----DEEDGEGGATLAFHFCMELRGPEPLGSRQRLIPMAAG 78
DB 27 ROYDGD-NSHVEKMLADEENADNMKA-----SVKRRFRNGR-----LCFAA-- 71
QY 79 RRAAPYLVTALLFTGAFLLGYVA-FRSGCOAGDS-----VLVYSEDVNEPDLDFH-- 132
DB 71 -----IALVIF--FLIGFMGYIGYCKRVQKECKYKLAETEEDTSMTETEDV 118
QY 132 --QGRLYWSIQAMFLQFEGREEDTIRQTS--LREVRVSGAMALTDIRALRSQ 186
DB 119 PTSSRLYADLKTLLSEKLNSTIEFADTIKQLSQNTYTPREGSQKDESLAYIENQFHEF 178
QY 187 KLDHWMTDHYVGLQFPRAHPNTLHWDEAKVGEQLPLEDPDYCYSAIGNTGLY 246
DB 179 KFSKVRKDEHYKIOVKSIGSQNMVTIVOSNGND--PVEPEGYVAFSKPEVSGKLY 235
QY 247 YAHGPRPELDLDRARGVDPVGRLLVVRGVISFAQKTYNADQFAGVLLYPPADPSQ 306
DB 236 HANGTCKDFEEL---SISVNGSLYIVRAGETTFEAKVANAQSFNAIGVLLYMDNKR-- 291
QY 307 DPPKPSLSSQAAVYGVHVLGTGDPYTPGFPSENOTQFPVASSGGLPSIPAOPISADIASR 366
DB 291 ---PVEADLALFGHAHLGTGDPYTPGFPSEHNTQFPSSQSGLPNIPVQISRAAEK 346
QY 367 LLARKIGVAAQEOGSLGSPYHLGPGPRRLVLYNNHRTSPINNITGCLIEGRSEPDHY 426
DB 347 LFGKKEGS-CPARWNID--SSCKLELSQNONMYLIVKNYLKERRLINIGVILGYDEPRY 404
QY 427 VVIGAQRAMPQG-AAKSAVGTAILELVRTFSSMWS-NGFPRRSLLFISMDGDFGSV 484
DB 405 VYVGAQRALAGVAAKSSVGTGLLKAQVFSMDISKDGFPPSSIIIPASTAGDFGAY 464
QY 485 GSTEMLEGYLSVHLKAAVYVSLDNAVLGDDKFNKTSPLTSLIESLYKOVDSPNHSQ 544
DB 465 GATEWLEGYLSLHLKAFYIYLDKVLGTSNFKVASAPLTYLGMKIMQDKHP-VDGK 523
QY 545 TLIEVYVNTNSMDAEVIRPLPMDSSAIFSFAFGVPAVESFEMDDQAPPLHAKEDTY 604
DB 524 SLYR-----DSNMISKV- EKLSFDNAAYFLAYSGIIPAVSFECEDAD--YPLTGRDLTY 576
QY 605 ENL-HKVLQGRPLPAVAQVVAOLAGOLLRLSHDRLLPLDFGQYGVVLRHIGNLEFSD 663
DB 577 EALIQVQO--LQNVKRAAEVAGOLIKLITDVELINDIYENKSLKFMKDLQKFTD 634
QY 664 LKARGLTIQWYSARGDYIRAEKLRQELISEERDELLTRYNVIRIMVEFYPLSQYS 723
DB 635 IRDMGLSIQWLYSARGDYFRATSRLLTDFHNAEKTRNFVMEIRIMKVEYHFLSPYVS 694

```

```

QY 724 PADSPFRHFWGRDHTLGLALDLRLRSNSGTPGATSTSGFOESFRROLATMTL 783
DB 695 PRESPFRHFWGSGSHLSALVENLKLROKN-----ITANETFLFRQALATMTI 745
QY 784 QGANALSGDVWNIDNNE 801
DB 746 QGANALSGDIWNIDNNE 763

RESULT 4
ID 099376 PRELIMINARY: PRT: 622 AA.
AC 099376;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE TRANSFERRIN RECEPTOR PROTEIN (TR) (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE: 91125359.
RA ROBERTS K.P.; GRISWOLD M.D.;
RT "Characterization of rat transferrin receptor cDNA: the regulation of
RT transferrin receptor mRNA in testes and in Sertoli cells in
RT culture."
RU Mol. Endocrinol. 4:531-542(1990).
CC -1- FUNCTION: CELLULAR UPTAKE OF IRON OCCURS VIA RECEPTOR-MEDIATED
CC ENDOCYTOSIS OF LIGAND-OCCUPIED TRANSFERRIN RECEPTOR.
CC -1- SUBUNIT: DIMER OF SIMILAR OR IDENTICAL CHAINS LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SIMILARITY: 88% SIMILARITY TO HUMAN TRANSFERRIN RECEPTOR
(CAC P02786).
DR EMBL; M58040; AAA42273.1;
KW Transmembrane; Glycoprotein; Receptor.
FT NON_TER 1
FT DOMAIN <1 622
FT FT EXTRACELLULAR
FT FT (CONTAINS TRANSFERRIN BINDING SITE)
FT FT (BY SIMILARITY).
FT CARBOHYD 109 109 POTENTIAL.
FT CARBOHYD 113 113 POTENTIAL.
FT CARBOHYD 179 179 POTENTIAL.
FT CARBOHYD 584 584 POTENTIAL.
FT CARBOHYD 589 589 POTENTIAL.
SC SEQUENCE 622 AA; 70152 MW; 7120562C CRC32;

Query Match 33.9%; Score 1424.5; DB 11; Length 622;
Best Local Similarity 46.5%; Pred. No. 5e-103;
Matches 304; Conservative 98; Mismatches 209; Indels 43; Gaps 15;

QY 155 DTIRTS---LREVRVSGAMALTDIRALRSQKLDHWMTDHYVGLQFPRAHPNTL 211
DB 5 DIKQLSQNTYTPREGSQKDENLAYIENT.FHDFKSKVRKDEHYKTIQVKNYSYQO-L 63
QY 212 HWVDEAGRVGEQLPLEDDVYCPYSALNVTGELYVAHYGPRPELDLDRAGVDPVGRLL 271
DB 64 VTINGSNID--PVEAEGVAVTSKAGEVYGLVHANFGKRDKEEL---NYSVNGSLV 117
QY 272 LVRRGVISFAQKVTNAQDFGAGVLIYEPADFSQDPKPSLSSQAAVYGVHVLGTGDPY 331
DB 118 IVRAGKITFAEKVANAQSFNAIGVLIYMDRNTF-----PVEADLQFFGHAHLGTGDPY 171
QY 332 TPGPSEFNOQFPVASSGLPSIPAOPISADIASRLKKGVPARQWQSGSLGSPYHL 391
DB 172 TPGPSEFNTQFPSSQGLPSIPVOTISRPAEKLFRNMGN-CPSWNID-SSCKLEL 229
QY 392 GPGPRRLVLYNNHRTSPINNIFGCLIEGRSEPDHYVYVIGAQRDAMPGAASAVGTAIL 451
DB 230 SQONVVKLTIVNNVILKETILNIFGVYKYEEDPRYIVVGAQRDAMGPGVAKSVGTGLL 289

```

| QY | 452 | ELVTFVSSVWS -NGRRPRRLSLFISMDGDDFGSSTEWLEGLYSLHKAVVYSLDNA | 510 |
|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|-----|
| Db | 290 | KLAQVFSMDISKDGRRPSRSTIFASHTACDYAVAGVETWLEGLSLHKAFATYINLMDV | 349 |
| QY | 511 | VLGDBKFRAKTSPLLTSLIESVLYKQVDSFNHSGQTLIEGVYFTFNSWDAEVRPLPMDS | 570 |
| Db | 350 | VLGTSNFKVASAPLLYTLMGKIMQVKKH -IDGKIYLR-----NSMWIK -IEELSLDNA | 402 |
| QY | 571 | AVSFPAENVGPAVFEFSFEMDDQAYPLHTKEDTYENLHKVLOGRLPVAQAQV---AQLAG | 627 |
| Db | 403 | APFLYASGIPAVSFCEDEDED-YPLGTRKLTLYE---ILIOKVQOLNQMRATAEAVG | 457 |
| QY | 628 | OLLILSHRLPLDPCRGVDVLRHIGNLNENSGDKLARKGLTQIMVYSARDYIRAAK | 687 |
| Db | 458 | QTLILTHDIELTLDLDMNSKLSLSPKLNQFKADIKMGSLTQVLAASRDYFPAISR | 517 |
| QY | 688 | LHQETIYSSSEDERLRTMIVNVRIMRVEFFLYSQVSPADSPRHIFMGRCGDTLGALDH | 747 |
| Db | 518 | LTTDDHNAKTRFVWRELNDRIKMEVHFSLPYSPRSRPHITFWGSGHTLSLVN | 577 |
| QY | 748 | LRLNSNSGTPGATVSTGFESEFRFRQRLATLWTLOGAANALSGDVWNIDNFF | 801 |
| Db | 578 | LRLRKN-----ITAFNETLPRNQLATLWTIOGVANALSGDIWNIDNEF | 622 |
| RESULT | 5 | | |
| Q90997 | | PRELIMINARY: PRT: 778 AA. | |
| ID | Q90997; | | |
| AC | 01-NOV-1996 (TREMBlrel. 01, Created) | | |
| DT | 01-NOV-1996 (TREMBlrel. 01, Last sequence update) | | |
| DT | 01-AUG-1998 (TREMBlrel. 07, Last annotation update) | | |
| DE | TRANSFERIN RECEPTOR. | | |
| OS | Gallus gallus (Chicken). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus. | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | GERHARDT E.M., CHAN L.N.L.; | | |
| RL | Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL: X55348; CAA39035.1; - | | |
| SO | SEQUENCE 778 AA: 86075 MW: 8895FP196 CRC32; | | |
| Query Match | 32.1%; Score 1349.5; DB 13; Length 778; | | |
| Best Local Similarity | 39.0%; Pred. NO. 5.2e-97; | | |
| Matches | 309: Conservative 134: Mismatches 267; Indels 83; Gaps 20; | | |
| QY | 21 | TYVQVVEGRKCHLEEF-EDGEDEGAETLAHFCPELKGPEPLGSRPROPNTLIPMAAGR | 79 |
| Db | 23 | STAROTDGD-NSHVEKMLKSADDEDEG-----DIERPEHMHVSMAPQ-----RNKG | 67 |
| QY | 80 | RAAPVLTUALLIFGAFLLGYVAFRGSQ-----ACGD-----SLVY-----VSE | 120 |
| Db | 68 | RLC-FLVIAAVLTLIGLIFGLYSTRGRIELARCODSGGCEITPTFASLYLVDGEGTVEE | 126 |
| QY | 121 | DVNYEPDDLDHOGRLVWSDLOAMFLOFLGEGRLIEDTIRQTSLRERYASGAGMAALTDIR | 180 |
| Db | 127 | ETGQPP-----VTFWPELKAAMLSKKSANKLVYD-----NLRMRGVDSFEGAEADTIN | 174 |
| QY | 181 | AAL-----SRQKLDHVWIDTYHYGLQFDPAPHPNTLHWYDEKAGVGEOLPLEDDPYCY | 235 |
| Db | 175 | MATYTHIEERFNLDFKVMWNEDEHYIKLO---VGSSTKNQVSISSINGKEET-LETPALIVAY | 229 |
| QY | 236 | SATIGNVTGELVYAHNGRPEDLQDRLARQGVDPVGRLLIVAVGISFAQKTYTNQDDGAGCV | 295 |
| Db | 230 | SSGSGVSGEPVYVNTGLKDFEITIQVVASLNGTIVYRAGKITTAEKANNAKEGAAAV | 289 |
| QY | 296 | LIIPEPADFSQDPKPRSLSSQOAVYGVHVALGTGDPYTPGFPFNOQFPVVASGGLPSIP | 355 |
| Db | 290 | LMYVDSLMY-----GITDILIPGCHALGIGCDPTTPFPSPFNHTQFPVYSSGGLPHIA | 342 |
| QY | 356 | ADPISADIASRLRLKLGKGVADQEWOGSLGSPYHLGCPRL -RLVYNNHRTSTPDIINI | 413 |

| | | | |
|--------|-----------------------------------------------------------------------|-----------------------------------------------------------------|--------------|
| Db | 343 | VQTISSSAALFESKMODTCEGKKAHISCKYTTQEOSEQINVKLDVYNSMKDKRLNTI | 402 |
| Qy | 414 | FGCIEGRSEPHYVYVIGAGORDAMPGAAKSAVGTAILLELVRTFSSMYSN-GERPRSSL | 472 |
| Db | 403 | FGALIOGFEEPRYVYVIGAGORDSMGPGVAKAGTGTAILLLELARIYISDIYKNKGYPKRSII | 462 |
| Qy | 473 | FISMDGDFGVSGETEMLEGLYSLVHLKAVVYVSLDNAVLGDDKFFAKTSPLLTSLIESV | 532 |
| Db | 463 | FASWSAGCYGAVGATEWLEEGYSAMLHKAFFYISLDAPVYGASHVKISASPLLMLLGSII | 522 |
| Qy | 533 | LKQVDSPPHSQOLTEYEOGVFFNPSMDAENVIRPLPMSDSAVSFTAFVVPVPEESFEMDDQ | 592 |
| Db | 523 | MKGAVNPAAVSESISYINRL--GPMWKAAY-PLGIDNAAEPPLAISOIPVLSFGFIYKDE | 578 |
| Qy | 593 | APPELTKEDYENYLHKYLOGRLPAVQAQAOLGQLIILFSLDRLPLDFGCHYGDVLR | 652 |
| Db | 579 | EYRFELDTGDTLEMLRKI--DNLDALLAAAEVAGQAALTLTHDHELFDIGRSEELLA | 636 |
| Qy | 653 | HIGLNIEFSGDLKARGLTLQWVYASRGDIYPAEKLROEIIYSSEERDLTRYNVRIMR | 712 |
| Qy | 713 | VEYFELSGYVSPASPFRHITMGCDITLQALDHLTLKLSNSSGTGATSSIGFOESRF | 772 |
| Db | 697 | VEYDPLSYLSPKDVPRPHIFGKGPHTLSLVEHLQLLXTNRSV-----DLNL | 747 |
| Qy | 773 | RRQALLTWTLQ 785 | |
| Db | 748 | REGQALATWTIKG 760 | |
| RESULT | 6 | | |
| P70627 | | | |
| ID | P70627 | PRELIMINARY; | PRT; 752 AA. |
| AC | P70627 | | |
| DT | 01-FEB-1997 | (TREMblrel. 02, Created) | |
| DT | 01-FEB-1997 | (TREMblrel. 02, Last sequence update) | |
| DT | 01-NOV-1998 | (TREMblrel. 08, Last annotation update) | |
| DE | NAAG-PEPTIDASE (GLUTAMATE CARBOXYPEPTIDASE II) (EC 3.4.17.21). | | |
| OS | Rattus norvegicus (Rat). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | |
| OC | Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. | | |
| NC | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=SPRAQUE DAWLEY; TISSUE=BRAIN; | | |
| RX | MEIDINE: 98041505. | | |
| RA | BZEDEGA T., TURTI T., WROBLEWSKA B., SHE D., CHUNG H.S., KIM H., | | |
| RA | NEALE J.H.; | | |
| RT | "Molecular cloning of a peptidase against N-acetylaspartylglutamate | | |
| RT | from a rat hippocampal cDNA library." | | |
| RL | J. Neurochem. 69:2270-2277(1997). | | |
| RN | [2] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TISSUE=BRAIN; | | |
| RC | MEIDINE: 98169524. | | |
| RA | LUTHI-CARER R., BERGER U.V., BARCZAK A.K., ENNA M., COYLE J.T.; | | |
| RT | "Isolation and expression of a rat brain cDNA encoding glutamate | | |
| RT | carboxypeptidase II." | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 95:3215-3220(1998). | | |
| RN | [3] | | |
| RP | SEQUENCE OF 284-752 FROM N.A. | | |
| RC | TISSUE=BRAIN; | | |
| RC | MEIDINE: 96149377. | | |
| RA | CARTER R.E., FELDMAN A.R., COYLE J.T.; | | |
| RT | "Prostate-specific membrane antigen is a hydrolase with substrate and | | |
| RT | pharmacologic characteristics of a neuropetidease." | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 93:749-753(1996). | | |
| RN | [4] | | |
| RP | SEQUENCE OF 284-752 FROM N.A. | | |
| RC | TISSUE=BRAIN; | | |
| RA | LUTHI-CARER R., BERGER U.V., BARCZAK A.K., ENNA M., COYLE J.T.; | | |
| RL | Submitted (DEC-1997) to the EMBL/GenBank/DBD databases. | | |

RESULT 8
C:3748
ID 043748 PRELIMINARY: PRT: 750 AA.
AC 043748:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE PROSTATE-SPECIFIC MEMBRANE ANTIGEN.
GN PSM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA O'KEEFE D.S., SU S.L., LUO Y.L., HORIGUCHI Y., BACICH D.J.,
RA POMEIL C.T., ZANDVILLET D., RUSSELL P.J., MOLLOY P.L., NOMAK N.U.,
RA MULLENS C., VONDER HAAR R.A., FAIR W.R., HESTON W.D.W.,
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF007544; AAC83972.1; -
SQ SEQUENCE 750 AA; 84304 MW; 8371DCB9 CRC32;

Query Match 17.3%; Score 727; DB 4; Length 750;
Best Local Similarity 27.3%; Pred. No. 2.1e-48;
Matches 213; Conservative 142; Mismatches 323; Indels 102; Gaps 22;

QY 65 RRGQNLIPMAAGRRAPYLVTALLITGAFILGYVAFRSCQACGSDVLVSEDEVY 124
DB 17 RRR-----WLCAG-----ALVLAGGFLLGLF-----FGFIRKSSNEATNI 52
QY 125 EPPDLFHOGRLYMSDLOAMFLOFLGEGRLIEDTIRQTSLEERVASGAMALTQDIRAALS 184
DB 53 TRK-----HMKAFILDELKA-----ENIKRFLHNFQIOPHLAGTEBONQKLAKOIOSQK 101
QY 185 ROKLDHWTDTHYVGLQFPDPAHPNTLHWVDAGKVEQLPDED-----DYCCPY 235
DB 102 EFGDLSVELAHYDVLSTYENKTHPNYISIIENDEGNEIFTSLEPPPGYENWDIVPEF 161
QY 236 SAI-----GNVTGELVYAHYGRPEDLODL-RARGVDPRGLLVYRGVYSIAQCVTAADQG 291
DB 162 SAFPSQGMPEGDLVYVNTAKTEDFFKLEKDMKINCSGKIYARGVPERGNVKNNAQLAG 221
QY 292 AGCVLIYEPEDFSDPPKPSLSQAAVYGVHL-----GTDPYTPGPFSPNOTQFP 344
DB 222 AGCVLISDPADYFAPGVKSYDPDGNMLPGCGVQRCNIIILNAGADPLTPGYANETAYNR 281
QY 345 PVASS-GLPSTIAPQISADIASRLRLKLGVPAP-OEWQSLGSPYHLPG-----P 395
DB 282 GIAEAVGLPSIPVHPIGYDAQKLEKMGSAAPDSSWNGS-LKVPYVNGPFGTGNFTQ 340
QY 396 RLRLVYNNHRTSTPINNIFGCIEGSRSEPHYVYVIGAQRAMGPGAAKSAVGTAILLELVR 455
DB 341 KKKMHISHNEVTRITVNIIGTLRGAVEPRYVILGHRDSWFGIDPOSGAAVVAHEIVR 400
QY 456 TFSWSNMGFRRRSLFLFISMDGDFGSVSTEWLEGYSLVHLKAVYVSDNNAVLGDD 515
DB 401 SFGTLKKEGMRRRRTILFASWDAEFGILGSTMAEENRLLQEGVATINADSSLEGYV 460
QY 516 KFHATSPLLTSLIESVLKQVDSPNH--SGQTLYEQVVFNTPSMDAEVIRPLPMDSSAYS 573
DB 461 TLRVCTPLMYSLVHNLTKELSKSPDEGFGKSLYESWTKKSPSPSPGMRISKLGSND 520
QY 574 FNAF---VGVPAVEFSFEMDDQ-----AYPLHTKEDYTEENLHKYLOGLAPVAQVAQL 625
DB 521 FVFVFORGLIASGARAYTKNMETNKFSGPLSHVYEIYELVEKEVDPMF-KYHLTAAVY 579
QY 626 AGQLIRLSHDLPLDLDFRGYGVVLRHIGNLNEFS---GDLKARGDLQWVYSARGDY 681
DB 580 RCGWVFEELANSYVLPDCCDYAVVLRKYADKIYISMKRPQEKITYSVSFDLSFSAVKNF 639
QY 682 IBAEKRLQEIYSSSEDERLRTMYNVRIMRVEFFYLSQVSPADSPF-RHIFMGSGDHT 740
DB 640 TEIASKFSRLQDPKSNPIVLRMMNDQLMFLERAFIDELGLP-DRPFYRHVITYAPSSHN 698

QY 741 LGALLDHLRLRSSNSGTPGATSTSGFOESR-----FRROLALTWTWLOGAANALS 791
DB 699 KYA-----GSEFPGLYDALFDIESKVDPSKAMGEVYKROIYAAFTVQAAAEITLS 747

RESULT 9
C35409
ID 035409 PRELIMINARY: PRT: 752 AA.
AC 035409:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE PROSTATE-SPECIFIC MEMBRANE ANTIGEN HOMOLOG.
GN MORSM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-SWISS WEBSTER NUDE;
RA BACICH D.J., HESTON W.D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF026380; AAB81971.1; -
SQ SEQUENCE 752 AA; 84635 MW; 868DAB06 CRC32;

Query Match 17.28%; Score 722; DB 11; Length 752;
Best Local Similarity 27.0%; Pred. No. 5.3e-48;
Matches 217; Conservative 143; Mismatches 305; Indels 138; Gaps 28;

QY 60 EPLGRRPQPNLIPMAAGRRAPYLVTALLITGAFILGYV---AFRGSCQACGSDVL 116
DB 12 EYLGHRR-----WLRVG-----TLVALITGFLIFLGWFLKPSNEATGN--- 54
QY 117 VSEDEVNTEPDLDFHOGRLYMSDLOAMFLOFLGEGRLIEDTIRQTSLEERVASGAMALT 176
DB 54 -VS-----HSG-----MKKEFLHMKENIKKFLYNTTRPHLAGTQNNFELA 95
QY 177 QDIRAALSROKLDHWTDTHYVGLQFPDPAHPNTLHWVDAG-----KYGEQLP--LED 228
DB 96 KOIHDMKREFGIDLVELSHYDVLSTYENKTHPNYISIIENDEGNEIFKTSLEOQPPGYEN 155
QY 229 -PDVYCPYSAL---GNVTGELVYAHYGRPEDLODL-RARGVDPRGLLVYRGVYSIAQK 283
DB 156 ISDVYPTSAFSPGTPGDLVFNARTEDFFKLEKMKISCSGKIYARGVYGFYRGNM 215
QY 284 VTNADFGAQGVLIYEPEDF-----SODPPKPSLSQAAVYGVH-HL-GTGDYTPGFP 336
DB 216 VKNVQLAGAKMILYSDPADYFVGVKSYPDGWNMLPGCGVQRCNIIILNAGADPLTPGYR 275
QY 337 SFPNOTQPPVASS-GLPSTIAPQISADIASRLRLKLGVPAP-OEWQSLGSPYHLPG 394
DB 276 ANEHAYRHELTEAVGLPSIPVHPIGYDAQKLEKMGSAAPDSSWNGS-LKVPYVNGPG 334
QY 395 -----RLRLVYNNHRTSTPINNIFGCIEGSRSEPHYVYVIGAQRAMGPGAAKSAVGT 447
DB 335 FAGNFTOKVKKMHISHYKYVTRITVNIIGTLGALPEBDYVILGHRDWWFVGIDPOSGA 394
QY 448 AILLELVTFSSSNVNGFRRRSLFLFISMDGDFGSVSTEWLEGYSLVHLKAVVYVSL 507
DB 395 AVVHEIVASFGLTKKGRRRRTILFASWDAEFGILGSTMAEHSLLQERVAAYINA 454
QY 508 DNAVLGDDKFRAKTSPLLTSLIESVLKQVDSPNH--SGQTLYEQVVFNTPSMDAEVIRPL 565
DB 455 DSIIEGNTYTLRVCTPLMYSLVHNLTKELQSPDEGFGKSLYD-----SMKEK--SPS 505
QY 566 PMSDASYFTAFVGVPAVEFSFEMDD-----QAYPLHT 599
DB 506 P-----BFIQMPRIKSLGSGNDEVFVFORGLIASGARAYTKNMKTNRYSSPYLHS 556
QY 600 KEDYTEENLHKYLOGLAPVAQA--AVAQLAGQLIRLSHDLPLDLDFRGYGVVLRHIGNL 657

Db 557 VYETELVAFYD---PTEKYHLLTVAQVBRAMVELANSIVLPEDCQSAVALKKYADTI 613
QY 658 NESS-----GDLKARGTLQWVYSGRGYIYAAEKLREIYSSSEERDLTRMVRMRY 713
Db 614 YNISKHPOEMEMAMISFSLFSAVNNETVASKFNQRLQELDKSNILLRINNDQMLYL 673
QY 714 EYFELSQYVSPASPFPHIFMGRGDH-----TLGALLDLHLLRLRSNSGTPGATSTGTF 767
Db 674 ERKFIDPLGLGPRFPYRHIIYASSHKYGESPFGIYDLFISKVNASKA----- 727
QY 768 QESRFRRLALLTWTLQGAANAL 790
Db 727 -WSEVKROIATETVQAAETL 748
RESULT 10
ID 077564 PRELIMINARY; PRT: 751 AA.
AC 077564;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE FOLYLPOLY-GAMMA-GLUTAMATE CARBOXYPEPTIDASE.
US Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=JUNAL MUCOSA;
RX MEDLINE; 98352082.
RA HALSTED C.H., LING E., LUTHI-CARTER R., VILLANUEVA J.A., GARDNER J.M.,
RA COYLE J.T.;
RT "Folylpoly-gamma-glutamate carboxypeptidase from pig jejunum.
RT Molecular characterization and relation to glutamate carboxypeptidase
RT II.";
RL J. Biol. Chem. 273:20417-20424(1998).
DR EMBL; AF050502; AAC39269.1; .
KW Carboxypeptidase.
SQ SEQUENCE 751 AA; 84523 MW; 0F6E680A CRC32;

Query Match 16.9%; Score 709.5; DB 6; Length 751;
Best local Similarity 26.9%; Pred. No. 5e-47;
Matches 212; Conservative 144; Mismatches 314; Indels 117; Gaps 25;
QY 65 RPRQPLIPMAAAGRRAAPLVLTALLIFGAFLLGVAFRGSCQACGDSVLVYSEDVNY 124
Db 17 RPR-----WLCAG-----ALVLAAGLFVLGFL-----FGWFIKSPNEANL 52
QY 125 EPRLDHFQGLYMSDQAMFLQFGEERLEDTIRQTSLSREVRAGSAGMAALTODIRAL 164
Db 53 SP-----HNVKALIDELKAENITFTFYNFRIPLAGTEQNFOLARQIOSQWK 102
QY 185 ROKLDHWYTDTHYVGLQEPDPAPHNTELMWYDEAGKVGEOPLDEP-----DYVCPY 235
Db 103 EFLDSELHAYVLSYPRKTRPNYISIIDEGNEIFNTSLDEPPPGYENNSDVVPF 162
QY 236 SAI---GNVTGELVYAHYGRPEDLOD-RAAGVDPVGRLLLVRYGVISFAQVYNAODFG 291
Db 163 SAFSPQGMPEGLVYVYVYARTDEDFEKERDMKINCSGKILLIARCKIFFRGNKYKNQALAG 222
QY 292 AOGVLIYREPADF-----SDDPRPSLSSQOAYYGHV-HL-GTGDPTTGPSPSNOTOPF 344
Db 223 AKGIILYSDADYFAPGVOSYPRGWNLPGGGVORGNILNLNGADPLTPGYPANEVAYRL 282
QY 345 PVASS-GLPSIPAPQISADLASRLRLKLGKPVAPQF-WQGSLLGSPYHLGPG-----P 395
Db 283 QIAEAVGLRPIYVPHYPSDAOKILLEKMGSAAPDDSWKKS-LHVPYNVGGFTGNGNSTQ 341
QY 396 RLRLVNNHRTSPPIINIFCIEGRSEPDHYVYIGAQRDAMGPGAAKSAYGTALLLELVR 455
Db 342 KYKMHINSDKNKRYIVNVIGTLRGAVERPDRYVILGHRDSWVFEGIDPOGGAANVHEIVR 401

QY 456 TFSWVSNGRPRRSLFTSMOGDFGVSQSTWTEMLGYLSVLKXAVVYSLDNVAGDD 515
Db 402 SFGKLKKEGMRPRRYVLFASWDAEYGLFGSTWMAEENSRILQERVAATNADSIENY 461
QY 516 KFAKTSPLLTSLIESVLQVQDSPNH--SGQTLVEQVFTNPSPDAEIVRP-----PMD 569
Db 462 TLAVDCTPLMYSLVYVLTRELQSPDGEFKSLFE-----SWNEKSPSESGLPRI 514
QY 570 ---SAVSFAF---VGVPAVESFMEF-----DQAVPLHTKEDTYENLHKVLOGLP 618
Db 515 KLGSGNDVEFVFORLGIASGRARYTRKDWYTKFSSYPLXHSVYETVELVEKFYD---PTF 571
QY 619 AQ--AAQLAGOLLIRLSHRLPLDGRYGDVVLHIGLNFEF---GDLKARGTLTQ 672
Db 572 KYHLAAVQVGGIVFELANSVPRFCDRAVAVLRNYADKLXINISMHPQEMATYVSFD 631
QY 673 WYVSARGDYIRAAEKLROEYSSSEERDLTRMYNVRIMRVEPFYVQVYSPDSFPHI 732
Db 632 SLFSVKNFTEIANSFNSERQDLDKNNPILLRIMNQMLHEAFIVPLGPPRAYRYHV 691
QY 733 EMGRGDHTLQALLDLRLRSNSGTPGATSTGFOESR-----FRQALALLTWTL 783
Db 692 IYAPSSHN-----KMGESFPGIYDALFDIENKVDPSKANGEVYROIATIAFTV 740
QY 784 QGAANAL 790
Db 741 QAAAGTL 747

RESULT 11
ID P91406 PRELIMINARY; PRT: 751 AA.
AC P91406;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE SIMILAR TO HUMAN PROSTRATE-SPECIFIC MEMBRANE ANTIGEN.
GN R57.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEERS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIKKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N.G., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOULDMAN P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA FAVELLO T., RIKIN L., CHIAPELLI B.;
RN Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RN Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RN Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.
RA WHITE S.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN
RP
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEERS M.,
RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAKINS T., HILLIER L., JER M., JOHNSTON L.,
RA JONES M., KERSTAN J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORITMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,
RA THIERRY-MIES J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; 278417; CAB01688.1; -.
SQ SEQUENCE 1483 AA; 168164 MW; C1FB0EAB CRC32;

Query Match 12.1%; Score 510; DB 5; Length 1483;
Best Local Similarity 24.7%; Pred. No. 5,9e-31;
Matches 164; Conservative 124; Mismatches 287; Indels 88; Gaps 21;

QY 146 QLTGEGRLDTITQTSREVRVAGSAGMALITODIRALSRKLDHWTDTHTVYGLQFPDP 205
DB 69 QLTINVDIRNSHLTKKRPVAGTENLRVAEMIRDMITQGLENVAFENYVNLISYPM 128
QY 206 AHNNTLHWDE-----AGKVGEOPLDPP-----VYCPSAIGNVTGELYAHYGR 252
DB 129 TTPNLTILKENDLVEFTSTGRSVYIKKEQNPPLAEIOMLAYSAACTVGDYVYNNAN 188
QY 253 PEDLDQIRANGVDPVGRLLLVYGVISFAQKVTNAODFGAGVLIYEPEDFESODPEKPS 312
DB 189 PSQIEYTESGIDLKDKIFLARYSSNRGNIAOMAYKAKACLVYSDPQVASLGCPN 248
QY 313 -----LSSQAVYGVHILGTGDPYTPGPFSF-----NOTQFPYVASGLPSIPAP 359
DB 249 ELYGNTDKMPSNTVQMSGVYIGLDDPRTPAFSPISIGDLFEKTEBDLDEKKIPIPLPI 308
QY 360 SAQIASRLKTKKGRVAPQOMOGSLGSPYHLGPG-----ERLRYVN-NHRST----- 409
DB 309 TYVTPAQILFENMKGDVADNADFOGK-LNVTRYRGGLINNOKLRTVYHAENEERSVIYEL 367
QY 409 -----PINNIFGCIERSEPDHYVVIQAORDAMGPGAKSAVGTATILLELVRTESS 459
DB 368 ISRLRLNFKRIQNIIMGVIKKSQEPDKFVLVSNHYDANTYGAVDNPGSTSTLLEYSRLAKQ 427
QY 460 MVS-NGFRPRRSLLFTSMWGDGFGSVGSTEMLEGYLSVHLKAVYVSLDNVAVLGDDK 518
DB 428 YONQGMVIFARSILFAHMDAEYGLIGSTFEAEYRLQLMRAVAIVNMD-LIGNGQTL 486
QY 519 AKSPPLTSLIESVLYKVDSPN-----HSGQTLVEQVVFNPSPMDAVIYR---LPPDSS 570
DB 487 GLSNPTVAVNLSAANAVEQNPPTMEQGRKTYLDSKRYAPSKNNSTPHYORIPAGGS 546
QY 571 AY-SFTAFVGPVAYEF--SPMEDDOAYPFLTKEDTYENLHKVLOGRLPA--VAQAIAOL 625
DB 547 DHPFEPYLGIPYVFFITSSLDAPPTPLVHTIYETPIYLIENIMD---PGYKHKAIAAGM 603
QY 626 AGQLLRSLSHDRLLPLDFGRYGDVYLRIHGNLNEFSGDLKARGLLTLOWVYSARGDYIRAA 685
DB 604 FICILTKFESKILPYDLNLM-----DSIFEYLPKLBDR-LNKTLIMGTKRDYLLDA 656
QY 686 EK-----LRQETIYSEEDER-LTRM-----YVAKIMRYEYFFLSQIYSPADSPF 729
DB 657 OKQFKLLQKTVLESLIYQVRNYSKLELPFGSRVDINNRLIEFEKCFINPHGAIGNQPA 716
QY 730 RHI 732
II :

DB 717 RHV 719

RESULT 14
ID 008919 PRELIMINARY; PRT; 783 AA.
AC 008919;
DT 01-NOV-1996 (TREMREL. 01, Created)
DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
DE CHROMOSOME XVI READING FRAME ORF YPL176C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN
RP SEQUENCE FROM N.A.
RA BENES V., RECHMANN S., NENTWICH U., VOSS H., ANSORGE W.,
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN
RP
RX MEDLINE; 783 AA; 88755 MW; 52926CCF CRC32;
SQ SEQUENCE 783 AA; 88755 MW; 52926CCF CRC32;

Query Match 7.1%; Score 297.5; DB 3; Length 783;
Best Local Similarity 22.4%; Pred. No. 8,3e-15;
Matches 126; Conservative 107; Mismatches 258; Indels 71; Gaps 19;

QY 152 RLDDITRQTSRLRRVAGSAGMALITODIRALSRKLDHWTDTHTVYGLQFPDPAPNLT 211
DB 160 KIERDLEYITSMPMMSGTSDAAIRHYIKESFKNQIRLAGEEFMAYSYNGVN--SL 216
QY 212 HWVDEAGKVGEOPLDPPDYCPYSAIGNVTG-ELYVAHGREPDIODIRARGVDPVGR 270
DB 217 RYVSKDDETFDIPLEBEN-FNPMHNGQNLNIPYIYANKASLDDNASMODGSLNGDFI 275
QY 271 LLYRVGISFAQKVTNAODFGAGVLIYEPADFSODPPKPSLSSQAVYGVHILGTGDP 330
DB 276 LLYVGDYDYOQMLT-AOEGAKALIFISPPYQDNK-----VIOMKSVALLPQYGTGDA 328
QY 331 YTPGFSFNQTOPPPYVASSGLPSIPAPISADIASRLKRLK---PVAPQOMOGSLGS 387
DB 329 LTPWEGSIRDPIDATEAKCLPPIPIPIISANGDKIILISDTGVKFSNNLFSGLINDC 388
QY 388 PYHLGCPRLRLVYNNH-RTSTPINNIFGCIERSEPDHYVVIQAORDAMGPGAKSAV 446
DB 389 -----RLDLVQAIRRHHVHDIYVGTIESQAGRAIYIAAPRNSASTGTMTPSFG 440
QY 447 TAILLELVRTFSSVSN-GERPRRSLLFTSMWGDGFGSVGSTEMLEGYLSVHLKAVYV 505
DB 441 TVVLLSLIQLOEYMYKKFMDKPLRNITYIFSGSEFENAGATELMKRTBA--LKSEIY 498
QY 506 SLDNAVUG--DD--KFAKTSPLTSLIESVLYK-----QVSPNNSGQTLVEQVVF 555
DB 499 IIDVGOIGIMDDNNLEIOCHPLLVDFOKNMTSRKFNKVDVNVHOG----- 547
QY 556 SMDAEYLRPLPMDSSAYSFATFVGPVAYEFSPMEDDOAYPFLTKEDTYENLHKVLOGRL 645
DB 547 DMTPIYLAOGIPV-----AIISSPGVM-----NREH-ITTVEDKDFIDKLDK 591
QY 616 PA-VAQAVNLAQQLIRLSHRLPLDFGRYGDVYLRIHGNLNEFSGDLKARGLLTLOWV 674
DB 592 KGEVLSBIMLYLVKESLDELIDDPFIPFSISNYVDPLSTLTKDLQKCPD-----TVNFD 645
QY 675 YSARGDYIRAAEKLRQETIYSE 696
DB 646 EVELGTTLMENKLOFEKWKSE 667
RESULT 15
061560

ID 061560 PRELIMINARY; PRT: 277 AA.
AC 061560:
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-NOV-1998 (TRENBLREL. 08, Last annotation update)
DE TRANSFERRIN RECEPTOR (FRAGMENT).
GN TRFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85159078.
RA STEARNE P.A., PIETERSZ G.A., GODING J.W.;
RT "cDNA cloning of the murine transferrin receptor: sequence of
RT trans-membrane and adjacent regions."
RL J. Immunol. 134:3474-3479(1985).
DR EMBL: M29618; AAA37616.1; -
DR MGD: MGI:98822; TRFR.
FT NON_TER 1 1
FT NON_TER 277 277
SQ SEQUENCE 277 AA: 31207 MW: EBE0C8AE CRC32;

Query Match 6.5%; Score 271.5; DB 11; Length 277;
Best Local Similarity 27.9%; Pred. No. 1,7e-13;
Matches 88; Conservative 54; Mismatches 116; Indels 57; Gaps 13;

QY 24 QNVEGPRKGLHLE---EEEDDEGCAETLAHFCPELNGPEPLGSRPRQPNLIPWAAG 78
Db 3 ROVDGD-NSHVEWKLADEENADNMKA-----SVRKPFRNGR-----LCFAA-- 47
QY 79 RRAAPYLVTALLFTGAFLLGYVA-FRSCQACGDS---VLVSEDEVNYEPDLDFH-- 132
Db 47 -----IALVIR--FLIGMSGYLGCKRVEQKECVKLAETETDKSETMETEDV 94
QY 132 --QRLYMSDLQAMFLQFLGEGRLIEDTIR---QTSLEFRVAGSAGMALTDIDRALSRQ 186
Db 95 PTSSRLYVADLTLSLSEKLNSTEFADTIKQLQONTYTPREAGSQKDESLAYIENGFHEF 154
QY 187 KLDHWVTDTHYVGLQFPDPAHPNLTLMWVDAGKVGQPLEDDPDVYCPYSAIGNVTGLV 246
Db 155 KFSKWRDEHYKIQYKSSIGONMVTIYQSNGLD---PVESPEGYVAFSKPTEVSGKLY 211
QY 247 YAHYGRPELDQDRAGVPRGRLILYRGVYSFAQKVTNADFGAQGLIYPERADFSQ 306
Db 212 HANFGTKDPEEL--SYSVNGSLVIVRAGEITFAEKVANAQSFNAIGVLITMDKNKF-- 267
QY 307 DPPKPSLSQQAAYG 321
Db 267 ----PVENDLALFG 277

Search completed: January 13, 2000, 12:12:05
Job time: 103 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 13, 2000, 08:36:16 ; Search time 1456.15 Seconds

(without alignments)
6772.760 Million cell updates/sec

Title: US-09-358-755-2

Perfect score: 2877
Sequence: 1 ctgcacgctcgcaggaggga.....ctgatataataagccaanaa 2877

Scoring table: IDENTITY_NDC

Searched: 4089388 seqs, 1713965092 residues

Database : EST.*

Word size : 0

Number of hits that pass the threshold : 8178776

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gp_est1:*
21: gp_est2:*
22: gp_est3:*
23: gp_est4:*
24: gp_est5:*
25: gp_est6:*
26: gp_est7:*
27: gp_est8:*
28: gp_est9:*
29: gp_est10:*
30: gp_est11:*
31: gp_est12:*
32: gp_est13:*
33: gp_est14:*
34: gp_est15:*
35: gp_est16:*
36: gp_est17:*
37: gp_est18:*
38: gp_est19:*
39: gp_est20:*
40: gp_est21:*
41: gp_est22:*
42: gp_est23:*
43: gp_est24:*
44: gp_est25:*
45: gp_est26:*
46: gp_est27:*
47: gp_est28:*
48: gp_est29:*
49: gp_est30:*

50: gp_est31:*
51: gp_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gp_est33:*
60: gp_est34:*
61: gp_est35:*
62: gp_est36:*
63: gp_est37:*
64: gp_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gp_gss1:*
70: gp_gss2:*
71: gp_gss3:*
72: gp_gss4:*
73: em_gss1:*
74: em_gss2:*
75: em_gss3:*
76: em_gss4:*
77: gp_gss5:*
78: gp_gss6:*
79: gp_gss7:*
80: gp_gss8:*
81: gp_gss9:*
82: em_gss5:*
83: em_gss6:*
84: em_gss7:*
85: em_gss8:*
86: em_gss9:*
87: em_gss10:*
88: em_gss11:*
89: gp_gss10:*
90: gp_gss11:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----------|--------------------|
| 1 | 542.8 | 18.9 | 548 | 61 | AI814648 | AI814648 wj75f09.x |
| 2 | 520 | 18.1 | 536 | 38 | AA748421 | AA748421 ny01a04.s |
| 3 | 481.4 | 16.7 | 700 | 50 | AJ725643 | AJ725643 RNJ725643 |
| 4 | 443.6 | 15.4 | 504 | 27 | AA007639 | AA007639 zh98d08.r |
| 5 | 414.4 | 14.4 | 782 | 43 | AU035332 | AU035332 AU035332 |
| 6 | 393.2 | 13.7 | 452 | 23 | R99295 | R99295 yq71d03.s1 |
| 7 | 388 | 13.5 | 444 | 22 | R46386 | R46386 yq50d07.s1 |
| 8 | 382 | 13.3 | 425 | 24 | R49459 | R49459 yq68f02.s1 |
| 9 | 371.8 | 12.9 | 444 | 27 | AA007614 | AA007614 zh99d08.s |
| 10 | 368.8 | 12.8 | 617 | 41 | AI047358 | AI047358 ud65d02.y |
| 11 | 361.4 | 12.6 | 384 | 34 | AA505641 | AA505641 nh82a03.s |
| 12 | 351 | 12.2 | 388 | 37 | AA682374 | AA682374 zj86a08.s |
| 13 | 348.6 | 12.1 | 363 | 38 | AA809686 | AA809686 n299d04.s |
| 14 | 334.6 | 11.6 | 384 | 24 | H63101 | H63101 yf50h04.s1 |
| 15 | 332.4 | 11.6 | 366 | 24 | H70432 | H70432 yf03c12.s1 |
| 16 | 324.8 | 11.3 | 503 | 22 | R35943 | R35943 yq68f02.r1 |
| 17 | 318.2 | 11.1 | 341 | 37 | AA693893 | AA693893 z149d07.s |
| 18 | 299.6 | 10.4 | 367 | 26 | W89178 | W89178 zh74h11.s1 |
| 19 | 292.6 | 10.2 | 708 | 41 | AI046365 | AI046365 ud65d02.x |
| 20 | 273 | 9.5 | 506 | 49 | AI663104 | AI663104 uk22g09.y |
| 21 | 267.4 | 9.3 | 408 | 21 | T64983 | T64983 yd11e08.s1 |
| 22 | 266.2 | 9.3 | 323 | 21 | T98167 | T98167 yq56c10.s1 |

| | | | | | | |
|----|-------|-----|------|----|----------|--------------------|
| 23 | 252.4 | 8.8 | 581 | 35 | AA537969 | AA537969 vj36d09.r |
| 24 | 236.2 | 8.2 | 421 | 45 | AI386238 | AI386238 mm20c12.y |
| 25 | 235.2 | 8.2 | 264 | 21 | T98166 | T98166 y66c10.r1 |
| 26 | 225.8 | 7.8 | 465 | 21 | T72291 | T72291 yc68e08.r1 |
| 27 | 224 | 7.8 | 225 | 38 | AA776476 | AA776476 z172h01.s |
| 28 | 220.6 | 7.7 | 384 | 48 | AI551005 | AI551005 vj28f05.y |
| 29 | 220.4 | 7.7 | 884 | 51 | AU050369 | AU050369 AU050369 |
| 30 | 220.2 | 7.7 | 1000 | 51 | CI81942 | CI81942 CI81942 |
| 31 | 214.8 | 7.5 | 269 | 24 | HE3187 | HE3187 yf50h05.r1 |
| 32 | 211 | 7.3 | 270 | 26 | W89177 | W89177 zh74h11.r1 |
| 33 | 208.4 | 7.2 | 675 | 48 | AI596094 | AI596094 uk22g09.x |
| 34 | 203.2 | 7.1 | 655 | 44 | AI256087 | AI256087 u194h08.x |
| 35 | 203 | 7.1 | 207 | 22 | R29632 | R29632 F1-1153D.22 |
| 36 | 203 | 7.1 | 436 | 51 | AU050545 | AU050545 AU050545 |
| 37 | 199.2 | 6.9 | 234 | 34 | AA511579 | AA511579 vj28f05.r |
| 38 | 198 | 6.9 | 234 | 21 | T87554 | T87554 yd89h12.s1 |
| 39 | 190.8 | 6.6 | 687 | 45 | AI132639 | AI132639 mm20c12.x |
| 40 | 184.6 | 6.4 | 461 | 43 | AI195082 | AI195082 u160d07.x |
| 41 | 157.2 | 5.5 | 457 | 41 | AI019669 | AI019669 ua92d01.r |
| 42 | 136.2 | 4.7 | 458 | 64 | AL120908 | AL120908 DKRP2762p |
| 43 | 116.4 | 4.0 | 699 | 51 | CI81968 | CI81968 CI81968 |
| 44 | 112.4 | 3.9 | 187 | 28 | AA096819 | AA096819 mm20c12.r |
| 45 | 101.4 | 3.5 | 106 | 24 | H70431 | H70431 ys03c12.r1 |

ALIGNMENTS

RESULT 1
AI814648/c
LOCUS
DEFINITION
AI814648 548 bp mRNA
W75F09.X1 NCI CGAP.Lu19 Homo sapiens cDNA clone IMAGE:2408681 3'
similar to PR:075422 075422 TRANSFERIN-RECEPTOR2.; mRNA
sequence.

ACCESSION
AI814648
NID
95425863
VERSION
AI814648.1 GI:5425863
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
On Jun 5, 1998 this sequence version replaced gi:3187613.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrrp/image/image.html

FEATURES
SOURCE
Seq primer: -40UP from GIBCO
High quality sequence stop: 460.
Location/Qualifiers

1..348
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2408681"
/clone_lib="NCI-CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"
/dev_stage="adult"

/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dt) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 126 a 135 c 165 g 120 t 2 others
ORIGIN

Query Match 18.9%; Score 542.8; DB 61; Length 548;
Best Local Similarity 99.3%; Pred. No. 1,4e-97;
Matches 544; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

| | | | |
|----|------|--------------------------------------------------------------|------|
| QY | 2326 | ttccagagagacgcttccgagctcagctgacgtcctgctacctgagacgtgcaaggcca | 2385 |
| DB | 548 | ttccagagagacgcttccgagctcagctgacgtcctgctacctgagacgtgcaaggcca | 489 |
| QY | 2386 | gccaatgctttagcgggagatgctggaacattatatacaactctgaaggccctggagat | 2445 |
| DB | 488 | gccaatgctttagcgggagatgctggaacattatatacaactctgaaggccctggagat | 429 |
| QY | 2446 | cctacatcccgctcccccagtcacagagctccctgctcctgctgaatgattcaaggt | 2505 |
| DB | 428 | cctacatcccgctcccccagtcacagagctccctgctcctgctgaatgattcaaggt | 369 |
| QY | 2506 | caggagagtgctgctagagctcagctcctcagctgagctgaattctcctccctacaca | 2565 |
| DB | 368 | caggagagtgctgctagagctcagctcctcagctgagctgaattctcctccctacaca | 309 |
| QY | 2566 | tctctcacgagagccagacccagacagatataccacacaccccccagctgagctgag | 2625 |
| DB | 308 | tctctcacgagagccagacccagacagatataccacacaccccccagctgagctgag | 249 |
| QY | 2626 | ctgacccctaatgtgacggtcattactgctggttaactcagagagatgagatcccttaac | 2685 |
| DB | 248 | ctgacccctaatgtgacggtcattactgctggttaactcagagagatgagatcccttaac | 189 |
| QY | 2686 | agcccttcccttctcggggtctcctacatcagagacacccctgagggttgctgaag | 2745 |
| DB | 188 | agcccttcccttctcggggtctcctacatcagagacacccctgagggttgctgaag | 129 |
| QY | 2746 | ccctgagacctgagcagctcgttagtggagagatcgtgagcaccatagacctatggcc | 2805 |
| DB | 128 | ccctgagacctgagcagctcgttagtggagagatcgtgagcaccatagacctatggcc | 69 |
| QY | 2806 | aacaggtggtctgtgtgaaagggtggtgagttcaatatcaataacacacgtgatatc | 2865 |
| DB | 68 | aacaggtggtctgtgtgaaagggtggtgagttcaatatcaataacacacgtgatatc | 9 |
| QY | 2866 | aataagcc 2873 | |
| DB | 8 | aataagcc 1 | |

RESULT 2
AA748421/c
LOCUS
DEFINITION
AA748421 536 bp mRNA
NY01804.s1 NCI CGAP.GCB1 Homo sapiens cDNA clone IMAGE:1270446 3'
similar to SW:TRSR_HUMAN P02786 TRANSFERIN RECEPTOR PROTEIN ;,
mRNA sequence.
ACCESSION
AA748421
NID
92788379
VERSION
AA748421.1 GI:2788379
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 536)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 14, 1998 this sequence version replaced g1:1798396.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/bdrrp/image/image.html

Insert Length: 1516 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 385.
 Location/Qualifiers
 1..536
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="1"
 /clone_1id="NCI-CGAP_GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="vector: p773D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, 19D-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). CDNA synthesis was
 primed with a Not I - oligo(dt) primer
 15'-TGTTACCAATCTAAGTGGAGGCGCCCTCATTTTGTGTTTGTGTTT-
 3'. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p773 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 119 a 125 c 160 g 132 t
 ORIGIN

Query Match 18.1%; Score 520; DB 38; Length 536;
 Best Local Similarity 99.1%; Pred. No. 4,4e-93;
 Matches 523; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2350 cagctagccctgcctcacctgcagctcgaaggcagcgaacatctgcctgaaggcagctgc
 |||||||
 DB 536 CAGCTAGCCCTGCCTCACCTGCAGCGCTGCAAGGCCAGCAATGGCTTAGCGGGAGTGC 477
 OY 2410 tggacaattgaatacaactctgagccctgaggatccctcacatccctccccaagtc 2469
 |||||||
 DB 476 TGGACAATTGATTAACAACCTCTGAGGCCCTGGGATCCTCAATCCCGTCCCAAGTCA 417
 OY 2470 agagctcctcgcctcgcctgcctgaatgattcaaggctcagggaggtgctcagagtc 2529
 |||||||
 DB 416 AGAGCTCCTCTGCTCCGCTTGAATGATTCAGGAGGTGAGGAGGTGCTCAGAGTCCACC 357
 OY 2530 tctcatgctgatcaattctcattacccctcacacatctcccaaggagcccaagccca 2589
 |||||||
 DB 356 TCTCATTTGCTGATCAATTTCTCATTTACCCCTACACATTTCTCCAGGAGCCCAAGCCCA 297
 OY 2590 gcacagatatcacacacacccagccctgcaagtgaatgctgacacctaattgtagcaggt 2649
 |||||||
 DB 296 GCACAGATATCACACACACCCAGCCCTGCAGTGATGACCTTAATGTGACGGGTCAATAC 237

OY 2650 tgcgtgtaatacagagagtagatcacccttcataatcacagcccttccttcctgaggtcc 2709
 |||||||
 DB 236 TGTCGGTTAATCAGAGAGTGTGATCCTCTTAATCAGACGCCCTTCCCTTTCTGGGGTCC 177
 OY 2710 tccatacttagagagaccactctgggaaggtttgtctaagccctgggaacctgagccactgt 2769
 |||||||
 DB 176 TCCATACCTTAGAGACCACTCTGGAGAGTTGTCTAGAGCCCTTGGAGCTGGCAGCTCTGTT 117
 OY 2770 agtggagagatgcctgcgcaccataagccttaatggccaacaggtgtctgtgtgaaagg 2829
 |||||||
 DB 116 AGTGGAGAGATGCTGTGGCCACATAGCTTATGGCCCAACAGGTGTCTGTGTGAAGGG 57
 OY 2830 gcctgaggttcaatcatcataaaccacctgatatcataagccaata 2877
 |||||||
 DB 56 GCGTGGAGTTTCATATCAATTAACCACTGATATCAATTAATTAATAA 9

RESULT 3
 AJ225643
 LOCUS AJ225643 700 bp mRNA
 DEFINITION RNU225643 Rat liver ESTs (E.Olivier) Rattus norvegicus cDNA clone
 IRL413, mRNA sequence.
 ACCESSION AJ225643
 NID 94833473
 VERSION AJ225643.2 GI:4833473
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutharia; Rodentia; Sciurognathu; Muridae; Murinae; Rattus.
 1 (bases 1 to 700)
 AUTHORS Olivier E., Soury E., Risler J.L., Smith F., Schneider K.,
 Lochner K., Douzeau J.Y., Rey G., and Saller J.P.
 TITLE A novel set of hepatic mRNAs preferentially expressed during an
 acute inflammation in rat represents mostly intracellular proteins
 JOURNAL Genomics 57 (3), 352-364 (1999)
 MEDLINE 99263497
 COMMENT On May 6, 1998 this sequence version replaced g1:3114939.

CONTACT: E. Olivier
 U78 INSERM
 543 chemin de la Breteque, 76233 Bois-Guillaume, France
 library construction: J.B.C. vol.270, p29998-30006 (1995)
 POLY-A-No.

FEATURES
 source
 1..700
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone_1id="IRL413"
 /clone_1ib="Rat liver ESTs (E.Olivier)"
 /tissue_type="liver"
 /dev_stage="adult"
 /note="Organ: liver; Vector: lambda ZAP11; Library
 construction: Ripberger, J.A. et al. J. Biol. Chem. (1995)
 270(50):29998-30006"

BASE COUNT 149 a 198 c 210 g 143 t
 ORIGIN

Query Match 16.7%; Score 481.4; DB 50; Length 700;
 Best Local Similarity 84.8%; Pred. No. 2e-85;
 Matches 598; Conservative 0; Mismatches 101; Indels 6; Gaps 5;

OY 1464 tgaatttgaagcgtgggcctcaagaggtgctgaaggcttacctgaagcgtgtgaacct 1523
 |||||||
 DB 1 TGACTTTGGCAGCGTGGAGGACGAGAGTGG-1TGAAGGCTTACCTCAGCGTGTACACT 59
 OY 1524 caaagccgtagtgtacgtgagcctggaacacgagctgtggtggatgatcaagtttcatgc 1583
 |||||||
 DB 60 CAAAGCGTGTGTATGTAGCGCTGAGCAACCTCGTGTGGAGAGCGCAATTCATGTC 119
 OY 1584 caagaccagccctcttgacaagctcattgagagtgctcctgaagcagtgatctcc 1643

| | | | | |
|------------|------|----------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|----------------------|
| | DB | 339 | CCAGATGCCCCACAAAGCAGGCCTGTCTTAGCCACCAGAGCGTGTACGGACATGTGCACCT | 398 |
| OY | 1002 | gggaacttgagaaacccctcacacacgtgagtcctccttccttcaatacaaacaggcttcctc | 1061 | |
| DB | 339 | GGAAGCTGGAGACCTTGTCACACACTGGGCTTCCCGTCTTCATCAACAACCATTCCTTC | 458 | |
| OY | 1062 | agttgcattcatgaagccttcccagatccacagccccagaccatcaagtgcacagattgctc | 1121 | |
| DB | 459 | AGTAGAATCATCAGGCGCTTCAGCATGCCCGCCGACGCCCATTCAGTGCATGACATTGCTCA | 518 | |
| OY | 1122 | ccgcttgctgtagaaagctccaagaagccttgttgcccccacaagaatggcagggaagcctc | 1181 | |
| DB | 519 | TCAATTTGCTCAGAGAACATCACAGGCCCGGTGCTTCCACGAGNTGAAAGGCGACCTTN | 578 | |
| OY | 1182 | agggccccccttaacacccggcgccggccagcaagctggcgagtcagtgatgataaatcaca | 1241 | |
| DB | 579 | AGGCTCTCTCTTAATTGCTGGGACCTGGGGCCGACATTCAGCTGGAGGGCACAACCAACC | 638 | |
| OY | 1242 | ga--cttcacaccccatcaacaacatcttcgctgcctcatcgaagagc | 1285 | |
| DB | 639 | GAGCTGTTTACCCCATTAGTACATTTTGGCTGCATTGANGGCC | 684 | |
| RESULT | 6 | | | |
| R99295/c | | R99295 | 452 bp | mRNA EST 14-SEP-1995 |
| LOCUS | | yc71d03.sl Soares fetal liver spleen INFLS Homo sapiens cDNA clone | | |
| DEFINITION | | IMAGE:201221 3', mRNA sequence. | | |
| ACCESSION | | R99295 | | |
| NID | | g985896 | | |
| VERSION | | R99295.1 | GI:985896 | |
| KEYWORDS | | EST. | | |
| SOURCE | | human. | | |
| ORGANISM | | Homo sapiens | | |
| REFERENCE | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | |
| AUTHORS | | Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | |
| | | 1 (bases 1 to 452) | | |
| | | Hillier,L., Clark,N., Dubque,T., Elliston,K., Hawkins,M., | | |
| | | Holtman,M., Holtman,M., Kucaba,T., Le.M., Lennon,G., Marra,M., | | |
| | | Parsons,J., Rifkin,B., Rohlfing,T., Soares,M., Tan,F., | | |
| | | Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and | | |
| | | Wilson,R. | | |
| TITLE | | The WashU-Merck EST Project | | |
| JOURNAL | | Unpublished (1995) | | |
| COMMENT | | On Nov 29, 1993 this sequence version replaced gi:430583. | | |
| | | Contact: Wilson RK | | |
| | | Washington University School of Medicine | | |
| | | 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 | | |
| | | Tel: 314 286 1800 | | |
| | | Fax: 314 286 1810 | | |
| | | Email: est@wustl.wustl.edu | | |
| | | Insert Size: 1674 | | |
| | | High quality sequence stops: 262 | | |
| | | Source: IMAGE Consortium, LNL | | |
| | | This clone is available royalty-free through LNL; contact the | | |
| | | IMAGE Consortium (info@image.lnl.gov) for further information. | | |
| | | Insert Length: 1674 Std Error: 0.00 | | |
| | | Seq primer: Promega -21ml3 | | |
| | | High quality sequence stop: 262. | | |
| FEATURES | | Location/Qualifiers | | |
| SOURCE | | 1..452 | | |
| | | /organism="Homo sapiens" | | |
| | | /db_xref="GDB:370271" | | |
| | | /db_xref="taxon:9606" | | |
| | | /map="21" | | |
| | | /clone="IMAGE:201221" | | |
| | | /clone.lib="Soares fetal liver spleen INFLS" | | |
| | | /sex="male" | | |
| | | /dev_stage="20 week-post conception fetus" | | |
| | | /lab_host="DH10B (ampicillin resistant)" | | |
| | | /note="Organ: Liver and Spleen; Vector: pTZ193D (Pharmacia)" | | |

with a modified polylinker, Site_1: Pac I; Site_2: Eco RI, 1st strand cDNA was primed with a Pac I - oligo(dT) primer (5', AACTGAGAGATTATTAAGATCTTTTCTTTTCTTTT 3'), double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 106 a 102 c 135 g 102 t 7 others

Query Match 13.7% Score 393.2; DB 23; Length 452;
Best Local Similarity 96.4% Pred. No. 4.2e-68;
Matches 432; Conservative 0; Mismatches 12; Indels 4; Gaps 3;

2426 attcttgagccctgggagatccatcccgccccccag--tcaagctccctgct 2483
Db attcttgagccctgggagatccatcccgccccccag--tcaagctccctgct 2483
452 ACTTGTGAGGCCCTGGGAGACCTCAATCCCGTCCCCCAGTMAAAAGCTCTCTGCT 393
2484 cctgccttgatgattcagagtgagagtgctcagagctccatctcattgctgac 2543
Db cctgccttgatgattcagagtgagagtgctcagagctccatctcattgctgac 2543
392 CTTCCCTTGAATGATTCAGGCTCAGGAGTGGCTCAGATCCACTCTCATTTGCTGATC 333
2544 aattctcatcaccctacatctctcagagagccagaccagacagacatccsc 2603
Db aattctcatcaccctacatctctcagagagccagaccagacagacatccsc 2603
332 AATTCTCTTACCCCTTANACATCTCTCAGAGAGCCAGCCCAAGCAAGATATCCAG 273
2604 aaccccaagccctcagtgagtgacacctaagtacagctacatcgcgttaacag 2663
Db aaccccaagccctcagtgagtgacacctaagtacagctacatcgcgttaacag 2663
272 ANACCCCAAGCCCTGAGGTAGCTGACCTTAATGAGCGTCACTACTGCGTTATACG 213
2664 agatgaagcaccctcacaacacagccctt-cccccttctggggtccatcactagag 2722
Db agatgaagcaccctcacaacacagccctt-cccccttctggggtccatcactagag 2722
212 AGAGTAGCATTCCTTCAATCAACAGCCCTTCCCTTCTGGGAGCTCTCATATACAG 153
2723 accactctggagagtggttgtaagccctggagaccctgctgtaagtgagagagac 2782
Db accactctggagagtggttgtaagccctggagaccctgctgtaagtgagagagac 2782
152 ACCACTCTGGAGGTGTGCTAGGCCCTGGGAGCTGCTGTGTTAGGAGATC 93
2783 gctggcaccatagccttctcagcacaagctgctgctgtaagtgagagagagagac 2842
Db gctggcaccatagccttctcagcacaagctgctgctgtaagtgagagagagagac 2842
92 GCTGGACCATAGCCTTATGAGCAAGCTGTTGTGTGAAGGGCGTGGAGTTTCA 33
2843 atataataaa-ccacctgatatcaata 2869
Db atataataaa-ccacctgatatcaata 2869
32 ATATCATTAACCCACCTGATATCAATA 5

RESULT 7
R46386/c 444 bp mRNA EST 22-MAY-1995
LOCUS y950d07.s1 Soares infant brain INIB Homo sapiens cDNA clone
DEFINITION IMAGE:36272 3', mRNA sequence.
ACCESSION R46386
NID 9822384
VERSION R46386.1 GI:822384
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 444)
AUTHORS Hillier, L., Clark, N., Dubugue, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohldmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 10, 1995 this sequence version replaced gi:805783.

Contact: Wilson RK

Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 1937
High quality sequence stops: 282 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 1937 Std Error: 0.00
Seq primer: Promega -2im13
High quality sequence stop: 282.
Location/Qualifiers

FEATURES

Source
1..444
/organism="Homo sapiens"
/db_xref="GDB:408773"
/db_xref="taxon:9606"
/clone="IMAGE:36272"
/clone_lib="Soares infant brain INIB"
/sex="Female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: lafmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5':
AACTGAGAGATTATTAAGATCTTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 97 a 92 c 127 g 123 t 5 others

Query Match 13.5% Score 388; DB 22; Length 444;
Best Local Similarity 97.2% Pred. No. 4.4e-67;
Matches 413; Conservative 0; Mismatches 10; Indels 2; Gaps 2;

2454 ccccgccccagtcagaagctcctcctcctcctcctcctgtaag-attcaggctcaggag 2512
Db ccccgccccagtcagaagctcctcctcctcctcctcctgtaag-attcaggctcaggag 2512
444 CCCCGTCCCCCAATCA-ANCTCGCTGCTCCTCGCTGATGATGATGAGGTGAGGAG 386
2513 gtgctcagagtcacacctcattgctgataatttcattaccctcaccatctcc 2572
Db gtgctcagagtcacacctcattgctgataatttcattaccctcaccatctcc 2572
385 NTGGCTCAGAGTCCACCTCTCATTCGTGANCATTTCTCATTTACCCCTANACATCTCTC 326
2573 acggagcccaagcccaagcagatatacacaaccccaagccctcagtgatgctgagcc 2632
Db acggagcccaagcccaagcagatatacacaaccccaagccctcagtgatgctgagcc 2632
325 ACGGAGCCCAAGCCCAAGCAGATATCCAAACCCCAAGCCCTGAGTGAAGTGAAGCC 266
2633 taatgtgacggtcactgctggttaatacagaagtagcatccctcattacacagccct 2692
Db taatgtgacggtcactgctggttaatacagaagtagcatccctcattacacagccct 2692
265 TAATGTGACGGTCACTGCTGATTAATCAGAGATGATGATGATGATGATGATGATGAT 206
2693 tcccccttctgggtggtcccaatacttagagaccctcctggagaggtttgctgaagccctgg 2752
Db tcccccttctgggtggtcccaatacttagagaccctcctggagaggtttgctgaagccctgg 2752
205 TCCTCTTCTGGGGTCTTCACCTACCTAGAGACCACTCTGGAGAGTTCTCTGAGCCCTGG 146
2753 accgtgcaagctcgttgtagtgagagatcgctggcaccatagccttctgccaagagt 2812
Db accgtgcaagctcgttgtagtgagagatcgctggcaccatagccttctgccaagagt 2812
145 ACCTGCGCAAGCTCTGTATGTTGGAGAGATCCCTGGACCAATAGCTTATGGCCAAAGCT 86
2813 gttctgtgtgtaaaaggcgtgaggtttcaatatacaataaaacacactgatacaataagc 2872
Db gttctgtgtgtaaaaggcgtgaggtttcaatatacaataaaacacactgatacaataagc 2872
85 GGTCTGTGTGTAAGAGGGCGTGAAGTTTCATATCAATTAACCAACCTGATATCAATAAA 26
2873 caaaa 2877
Db 25 AAAAA 21

RESULT 8

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1734
High quality sequence stops: 326 Source: IMAGE Consortium, LINT
This clone is available royalty-free through LINT, contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1734 Std Error: 0.00
Seq Primer: Promega -21ml13
High quality sequence stop: 326.

| Query Match | 13.3% | Score 382 | DB 22 | Length 425 |
|-----------------------|----------------------------------------------------------------|-------------------|----------|------------|
| Best Local Similarity | 96.7% | Freq. No. 6.7e-66 | | |
| Matches 411 | Conservative 0 | Mismatches 11 | Indels 3 | Gaps 2 |
| QY 2455 | ccgcgcgccccagtcgaagagct-cctctgctctctgccttga--tgattcaggttcaggga | 2511 | | |
| Db 425 | ccggtcccccacatcagaagctccctctgctctcccttgaatganttacaggttcaggga | 366 | | |
| QY 2512 | ggttgctcagagtcacacctcatctgcttgatcaattctcattacccctacacatcttc | 2571 | | |
| Db 365 | ggttgctcagagtcacacctcatctgcttgatcaattctcattacccctacacatcttc | 306 | | |
| QY 2572 | caaggagcccgagaccgccgacagatattccacaacccagccctgcgcaagtgaagtgacc | 2631 | | |
| Db 305 | caaggagcccgagaccgccgacagatattccacaacccagccctgcgcaagtgaagtgacc | 246 | | |

| RESULT | 9 | | | | | | | | |
|------------|------------------------------------------------------------------|--------|------|-----|-------------|--|--|--|--|
| AA007614/c | | | | | | | | | |
| LOCUS | AA007614 | 444 bp | mrna | EST | 09-MAY-1997 | | | | |
| DEFINITION | z1999408.s1 Soares fetal liver spleen INFIL S1 Homo sapiens cdna | | | | | | | | |
| | clone IMAGE:429423 3' similar to PIR:AA8592 AA8592 transferrin | | | | | | | | |
| | receptor protein - Chinese hamster ; ; mRNA sequence. | | | | | | | | |

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 266 1800
Tel.: 314 266 1800
Fax: 314 266 1810
Email: estewatson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 652 Std Error: 0.00
Seq primer: mob.RGA+ET
High quality sequence stop: 337.

```

FEATURES
source
location/Qualifiers
1..444
/organism="Homo sapiens"
/db_xref="GDB:132912"
/db_xref="taxon:9606"
/clone="IMAGE:429423"
/clone_lib="Soares_fetal_liver_spleen_1NPLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker. Site_1: Pac I; Site_2: Eco RI. This is a subtracted version of the original Soares fetal

```

liver spleen INF1S library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGATTAATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonafido."

BASE COUNT 74 a 131 c 139 g 89 t 11 others

ORIGIN

Query Match 12.9%; Score 371.8; DB 27; Length 444;
Best Local Similarity 91.9%; Pred. No. 7e-64;
Matches 408; Conservative 0; Mismatches 33; Indels 3; Gaps 2;

620 tgggctcgaattcccgagctcac--cccaacaccgacgtgctcgaagc 677
|||||
444 TGGGCTGCAATTCGGGATCCNGCTCACACCAACACTTGCTGGTGTGATGAGC 385
678 cgggaagtcggaagacagctgcgcctggaagacccttactctcccc-tacagcg 736
384 CGGGAAGTCGGAGAGAGAGTCCGCTGGAGGACCCTGACCTACTGCCCTTACAAGC 325
737 ccatacggaagctcagggagagctggtgtacgcccactacggcgccggaagaccctgc 796
324 NCATGGCAACTCAGGGAGAGCTGCTGTACGGCAGCTACGGGCGCCGGAAGACTGC 265
797 aggaacctggcgagcgagggcgctgcatcagtgccgcctgctgtgtgctgcgagcg 856
264 AGGACTCGCGGGCCAGGGGGCTGATCCAGTGGGNCCTGCTGTGTGGCNGTGGGG 205
857 tgatcagcttcgcccagaagtgaccacatgctcaggaactcgggcctaagaagtgtc 916
204 TGANTCAGTTCGCCAGAGAGTGACCAATGCTCAGGACTTCGGGCTCAAGAGTGTCA 145
917 tataccagagacagcaggaactctccagagaccacccaagagcctgtctcagcagc 976
144 TATACCCAGAGCCAGGAGCTTCTCNCAGAGCCACCCAGCAAGCCGTGCTCANNAGC 85
977 aggcagcttatgacatgtcaccttggaacttgagaccctacacacacgtgctccct 1036
84 AGGAGTGATGACATNTGCACCTGGGAGACTGAGAGACCCCTACACACTGCTCCCT 25
1037 ccttaatacaaccagctccctc 1060
24 CCTCAATCAACCCAGNTCCCTC 1

RESULT 10
LOCUS A1047358 617 bp mRNA EST 08-JUL-1998
DEFINITION u65d02.y1 Sugano mouse liver m1a Mus musculus cDNA clone IMAGE:1450755 5' similar to TR:Q99376 Q99376 TRANSEFERIN RECEPTOR PROTEIN; mRNA sequence.
ACCESSION A1047358
NID 93295645
VERSION A1047358.1 GI:3295645
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 617)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theisinger, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On May 8, 1995 this sequence version replaced gi:800961.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:924071
Seq primer: custom primer used
High quality sequence stop: 494.
Location/Qualifiers
1. 617
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1450755"
/clone_lib="Sugano mouse liver m1a"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII (CACTGTG); Site_2: DraIII (CAGCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTG, 3' site CAGCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTTAAGCTGCG and 3' end primer CGACTGTGAGCTGACACA."

BASE COUNT 121 a 177 c 186 g 133 t

ORIGIN

Query Match 12.8%; Score 368.8; DB 41; Length 617;
Best Local Similarity 74.0%; Pred. No. 3e-63;
Matches 534; Conservative 0; Mismatches 82; Indels 106; Gaps 1;

1333 tggggccagagcagcttaactccgtctgtggagcagctatactcctggagctgtgcg 1392
|||||
1 TGGGCCCCAGGAGCAGCCAACTGTGAGTGGGAGCTGCACCTCGTAGACTGTTCGG 60
1393 accttctcctatggtgagagacggttcgcggcccgcaagctcctcttcacagc 1452
|||||
61 ACCTTCTTCCATGTCAGCAATGGGTTCAAGACTCGAAGAAATCTTTTGTATCAGC 120
1453 tgggacggtgctgacttggaagcgttggtcctcagagagtgctcagaagctaccacg 1512
|||||
121 TGGGACGGAGGTGACTTGGCAGGCTGGGAGCCACAGAGTGGTGGAGGCTTACTCAGC 180
1513 gtgtcgaacctcaagccgtatgttaagtgaagccttggaacaaagcagctgtgggagtag 1572
|||||
181 GTGCTACACCTCAAGAGCTGTGTGTAGCTGAGCTCGA----- 219
1573 aagttcatgcaagacagagccctcttgcaagctcattcattgagagtgctcgaagcag 1632
219 ----- 219
1633 gtgattctcccaacacacagctgggcagactctctatgacaggtggttcaaccaatccc 1692
|||||
219 -----CAGACCTCTATGAAACAAGTGGGAGCTGACCCACACCC 254
1693 agctggagctgtgagtgatccggccctaccacagcagcagctgctactcctcaag 1752
|||||
255 AGCTGGGATGCTGAAGTATTCACGCCCTGCCATGAGACACAGATGATATTCCTTCA 314
1753 gcttttggagagccctgcgcgtctgagttcctcttcttgagagcagcagcagcctacc 1812
|||||

Db 315 GCCTTGGGGGGTCCAGCTGTGAGTTCCTTCATGAGATGATCGGGTGAACCA 374
 QY 1813 ttcttgacacaaaggaggaacttaagagaacctgcaagtgtcgcaaggccgctg 1872
 Db 375 TTCTGTGACAGAGAGAGGACATATGAGATCTGCACAGAGATCTGCGAGTCTGCTG 434
 QY 1873 cccgcgctggcccgccgctggccagctgcagggagccgctccctcaccgctcagccac 1932
 Db 435 CCGCCCTGTGCTCCAGGAGTGGCTCAGTCCGGGCCAGCTCTCATTCGACTGAGCCAC 494
 QY 1933 gatcgctgtgcgcccctcgaactgcgcgcgaagggagcgtcgtcctcgaagcagatcg 1992
 Db 495 GATCACTACTAGCGCGTAGACTTGCGCGCATATGAGAGCGTGGTTCACGACATCGGC 554
 QY 1993 aacctcagaagttctctgtggagacctaaaggcccgggcggtgacctgcaagtgtgtac 2052
 Db 555 AACCTCATAGTCTCTGTGGGACCTCAAGGAGCGCGGCTGACCTGCACTGAGGTGTAC 614
 QY 2053 tc 2054
 Db 615 TC 616

RESULT 11
 AA505641/c 384 bp mRNA EST 18-AUG-1997
 LOCUS nh82803.s1 NCI_CGAP_Br1.1 Homo sapiens cDNA clone IMAGE:964972 3'
 DEFINITION mRNA sequence.
 ACCESSION AA505641
 NID 92241778
 VERSION AA505641.1 GI:2241778
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 384)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 9, 1995 this sequence version replaced gi:802599.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www.bio.lnl.gov/bdrip/image/image.html

Insert Length: 2525 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 256.

FEATURES

Location/Qualifiers
 1..384
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:964972"
 /clone_1id="NCI_CGAP_Br1.1"
 /sex="female, pooled"
 /tissue.type="breast"
 /lab_host="DH10B"
 /note="Vector: pT773D-Pac (Pharmacia) with a modified
 polylinker: 1st strand cDNA was prepared from pooled bulk
 breast tumor tissue, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT773

vector. Library is not normalized. (The normalized
 version of this library is NCI_CGAP_Br2.) Library was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 85 a 86 c 108 g 105 t
 ORIGIN

Query Match 12.6%; Score 361.4; DB 34; Length 384;
 Best Local Similarity 98.4%; Pred. No. 7.5e-62;
 Matches 365; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2507 agggaggtgctgaagctacacctctcattgctgcatcaattctcattaccctacacat 2566
 Db 384 AGGAGGTGCTCAGATTCACCTCTCATTCGATCAATTTTCATTACCCCTACACAT 325
 QY 2567 ctctcagagagcccaacccagacagatattccacaccccaagccctgcaagttagc 2626
 Db 324 CTCGCACGGAGGCCACACCCACGACGATATCCACACCCACGCCCTGCAGGTAGC 265
 QY 2627 tgacctaatgtgacggtacactgtcgtgtaatacagagagtagcatccctcaatcaca 2686
 Db 264 TGACCTTAATGTGACGGGTACTACTGTGTTAATCAGAGATGACATCCCTCAATCACA 205
 QY 2687 gcccttccccctttctgtggtctctcattacatagagacccactctggaggttgctaacg 2746
 Db 204 GCCCTTCCCTTCTGTGGGTCTCTCCATACCTAGAGACCACTCTGGAGGTTGGCTAGGC 145
 QY 2747 ccttgagacctggcgaagctcgttagtgtagagagatcctgcaccataagcttagtgcga 2806
 Db 144 CCTGGAGACCTGGCCAGCTCTGTAGTGGAGAGATGCTGGACACATAGCCTTAGGCCA 85
 QY 2807 acaggtgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2866
 Db 84 ACAGGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 25
 QY 2867 ataagcccaaaa 2877
 Db 24 ATAAAAAAA 14

RESULT 12
 AA682374/c 388 bp mRNA EST 19-DEC-1997
 LOCUS zj86a08.s1 Soares_fetal_liver_spleen_INFIS_S1 Homo sapiens cDNA
 DEFINITION clone IMAGE:461750 3' similar to gp:ML1507 TRANSFERRIN RECEPTOR
 PROTEIN (HUMAN);, mRNA sequence.

ACCESSION AA682374
 NID 92686655
 VERSION AA682374.1 GI:2669655
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 388)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Maria, M.,
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
 WashU-NCI human EST Project
 JOURNAL Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1394174.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 380.

FEATURES
source

Location/Qualifiers
1. .368
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="8"
/clone="IMAGE:461750"
/clone_lib="Soares_fetal_liver_spleen.INFLS.S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p773d (Pharmacia) with a modified polylinker; Site.1: Pac I; Site.2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACTGGAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 65 a 108 c 129 g 86 t

ORIGIN

Query Match 12.1%; Score 351; DB 37; Length 388;
Best Local Similarity 97.3%; Pred. No. 8.4e-60;
Matches 357; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

737 ccacgcgaacgtcacgaggaagctgtgtacgcccactacgagggcgccgaagacttcg 796
|||||
388 CCATCGCAACGTCAACGGAGAGAGCTGTGTAGCCCACTACGGCGCGCGAGAGACTGC 329
|||||
797 aggacctgcygagcgaagggcgctgtgacccagctgctgcgctgcgctgtgctg 856
|||||
328 AGGACCTCGGGGCGAGGGGCGGTGATCCAGTGGCGCGCTGCTGCTGCTGCTGCTG 269
|||||
857 tgatcagcttcgcccgaagggcgctgtgacccagctgctgcgctgcgctgtgctg 916
|||||
268 TGATCAGCTTCGCCGAGAGGCTGACCAATGCTCAGGACTTCGGGCTCAAGAGTCTCA 209
|||||
917 tataccagagcagcagcagcttcctccagagccaccagcagcagcagcagcagcagc 976
|||||
208 TATACCAAGCCAGCGGACTTCTCCAGAGACCACCAAGCCAGCTGTCCAGGCACG 149
|||||
977 aggcagtgatgacatgtgacccctgagcagctgagacccctacacaccttgcttcctt 1036
|||||
148 AGGACGTGTATGAGCATGTGACACTGGAACTGGAGACCCCTACACACTGGCTTCCCT 89
|||||
1037 ccttcataaaccagcttcctccagctgtgacatcatcagccttcctccagatcccaagcc 1096
|||||
88 CCTTCATTAACCCAGTTCCCTCCAGTTGCATCATCAGGCTTCCAGCATCCACGCC 29
|||||
1097 agcccat 1103
|||||
28 GAGCCCT 22

RESULT 13
AA809686 363 bp mRNA EST 18-FEB-1998
LOCUS n299d04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:130351 3',
DEFINITION mRNA sequence.
ACCESSION AA809686
NID 92879092
VERSION AA809686.1 GI:2879092
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 363)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE
JOURNAL
COMMENT

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2286784.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1482 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 361.
Location/Qualifiers
1. .363
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="954B04; 954B04"
/clone="IMAGE:1303591"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: p773d-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAGTGGAGGCGCGCTCAATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 83 a 77 c 106 g 97 t

ORIGIN

Query Match 12.1%; Score 348.6; DB 38; Length 363;
Best Local Similarity 98.9%; Pred. No. 2.4e-59;
Matches 351; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

2523 gtccacatctcatctgctatcaattctcatcaccctacacatctccacgagccca 2582
|||||
363 GTCCACCTCTCATCTGCTGATCAATTTTCATATACCCCTACACATCTCTCCAGGCCCA 304
|||||
2583 gaacccagcacagatataccacacacccagccctgcagtgtagctgacccaatgtgacg 2642
|||||
303 GACCCAGACAGATATCCACACACCAGCCCTGCAATGATGACCTAATGTGACG 244
|||||
2643 gtccatctctggttaacagagagtagatccctcaatcacagcccttcctctct 2702
|||||
243 GTCATATCTGTGCTGTTAATCAGAGATAGCATCTTTCATATCAGACCCCTTCCCTTTCT 184
|||||
2703 ggggtccctccatcactagagacacactctggaggttttgtaagccctgagcactgcccag 2762
|||||
189 GGGGTCTCTCCATACCTAGAGACACTCTGGAGGTTTCTAGGCCCTGGAGACTGGCCAG 124
|||||
2763 cctctgttagtgagagatcgctgacacacaccccttagcgaacagtggtctggtg 2822
|||||
123 CTCTGTATGTGGAGAGATCGCTGGACCATATGACCTTATGGCCAACAGTGTGTGTGT 64
|||||
2823 gaaagggcggtgagattccaataataaacacactatataaagcccaaa 2877
|||||

Db 63 GAAAGGGCGTGGATTTCATATCAATAAACCCGATCATATAAAAAA 9

RESULT 14
LOCUS H63101/c 384 bp mRNA EST 11-OCT-1995
DEFINITION Y50H04.s1 Soares fetal liver spleen INF1.S Homo sapiens cDNA clone
(HUMAN);, mRNA sequence.

ACCESSION H63101
NID 91017902
VERSION H63101.1 GI:1017902
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 384)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Matra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
COMMENT On May 8, 1995 this sequence version replaced gi:799827.

TITLE
JOURNAL
COMMENT

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 656
High quality sequence stops: 272
Source: IMAGE Consortium, LNL.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 656 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 272.
Location/Qualifiers
1..384
/organism="Homo sapiens"
/db_xref="GDB:3177890"
/db_xref="taxon:9606"
/clone="IMAGE:208759"
/clone_lib="Soares fetal liver spleen INF1.S"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pRT73D (Pharmacia)
with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGCGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Patrina Bonaldo."

BASE COUNT 66 a 104 c 126 g 78 t 10 others

ORIGIN

Query Match 11.6%; Score 334.6; DB 24; Length 384;
Best local Similarity 94.5%; Pred. No. 1.4e-56;
Matches 362; Conservative 0; Mismatches 19; Indels 2; Gaps 2;

QY 711 ccctgacgtctactgccccctacagcgccacatcggaacgtacggaagctgtgtcgcg 770
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 383 CCCTGANGTCTACTGNCCTAAAGCGCNCATCGMAACGTCA-GGAGAGNCCTGGGTAGC 325
QY 771 ccactacggcgcccgaaacactgcagacactgcggcgccagggcgctgtgattccactg 830

Db 324 CNACTACGGGNGGCGCCGAACACCTGCAGACCTGCGGCCAGGGGCGTGGATCACTGGG 265
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 831 ccgctctgctgtgtgctgcgctgtggtgtg-atacagcttcgcccaagaagtgtaccatgtctc 889
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 264 CCGCTGTGCTGTGTGTGCGGGGTGGAATACNANTCGCCCGAAGAGTGACCAATGCTC 205
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 890 aggaacttcggggtcgaagagtgctatataccagaagccagcgagctctcccaagacc 949
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 204 AGGACTTCGGGCGCTCAAGAGAGTCTCATATACCAAGCCAGCAGGACTTCTCCAGAGCC 145
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 950 caccacaagccaagctgtctcaagccagcagcagtgatgtgaaatgtgcacctgtggaact 1009
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 144 CACCCAGCCAGCAGCTGTCCAGCAGCAGCAGTGTATGACATGTGCNCTGGGAATG 85
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 1010 gaagaccctacacacactgtgcttccttcctcaatacaaacacagctccctccagtgcat 1069
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 84 GAGACCCCTACACACCTGTGCTTCCCTTCTTCATATCAATCAAAACCGATTCCTCCAGTTGCAT 25
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 1070 catcagggcttcccaagcatcca 1092
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 24 CATCAGGCGCTTCCACATCCCA 2
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||

RESULT 15
LOCUS H70432/c 366 bp mRNA EST 24-OCT-1995
DEFINITION Y503C12.s1 Soares fetal liver spleen INF1.S Homo sapiens cDNA clone
IMAGE:213718 3', mRNA sequence.

ACCESSION H70432
NID 91040638
VERSION H70432.1 GI:1040638
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 366)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapellil,B.,
Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,M.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellendberg,K., Soares,M.B., Tan,F., Thierry-Mieg,D.,
Trevasakis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Matra,M.
Generation and analysis of 280,000 human expressed sequence tags
97044478
On Apr 14, 1993 this sequence version replaced gi:837502.

TITLE
JOURNAL
MEDLINE
COMMENT

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 801
High quality sequence stops: 269
Source: IMAGE Consortium, LNL.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 801 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 269.
Location/Qualifiers
1..366
/organism="Homo sapiens"
/db_xref="GDB:3179022"
/db_xref="taxon:9606"
/map="11"
/clone="IMAGE:213718"
/clone_lib="Soares fetal liver spleen INF1.S"
/sex="male"

FEATURES
source

```

/dev stage="20 week post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dN) primer
[5' AACGCGAAGATTAATTAAGAAGCTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fátima Bernaldo."

```

| Query Match | 11.6% | Score 332.4 | DB 24 | Length 366 |
|-----------------------|-------------------------------------------------------------------|-------------------|--------------|------------|
| Best Local Similarity | 98.6% | Pred. No. 3.8e-56 | | |
| Matches 355 | Conservative | 0 | Mismatches 3 | Indels 2 |
| | | | | Gaps 2 |
| QY 2511 | aggtgctcagatgcctaccctctcatcgtg-tgtatcaatttctcatlaccctacacatc | 2569 | | |
| Db 366 | AGGTGNNTCAGAGTCCACCTCTCATTTGGCTGATCAATTCTCATTAACCCCTANACATCTC | 307 | | |
| QY 2570 | tcacagagagcccaagaccccaagacagatataccacaaccccaagccctcagtgatctga | 2629 | | |
| Db 306 | TCCAGGGAGCCCGACCCCGACAGATATCCACACACCCCGCCCTCGAGTGATCA | 247 | | |
| QY 2630 | cacctaatgtacagagtcatactgtcggttaatcagaagagatagatcccttcataccagcc | 2689 | | |
| Db 246 | CCCTAATGTGACGGTCATACCTGCTGGTTATATAGAGATGAGATCCCTTCATTCACAGCC | 187 | | |
| QY 2690 | ccttcacccttcttggtgtcctccatcacctagagacacactctggagaggtttgttaagccct | 2749 | | |
| Db 186 | CCTTCCTCTTCTGGGGTCTCCATACCTAGAGACCACTCTGGAGGTTTGGTTAGGCCCT | 127 | | |
| QY 2750 | gggaacctggccaagctctgttagtgagagaagatcgcctgagacatagcccttatggtccaaca | 2809 | | |
| Db 126 | GGGACCTGGCCAGCCTCTTACTGTGGAGAGAACGCTGGACCATTA-CCTTAATGGCCAA | 68 | | |
| QY 2810 | ggttgtctggttgtaaaagggtgttaggttttaactatacaataaacaccctgtatcaata | 2869 | | |
| Db 67 | GGTGGTCTGTGGTGAAGGGGCGTGGAGTTTAAATTCATTAACCACTGTATCAATA | 8 | | |

Search completed: January 13, 2000, 11:10:38
Job time: 9262 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 13, 2000, 02:17:22 : Search time 1880.46 Seconds

(without alignments)
6541.900 Million cell updates/sec

Title: US-09-358-755-2

Perfect score: 2877

Sequence: 1 ctgcagcgtctcagagagga.....ctgatatacaataagccaana 2877

Scoring table: IDENTITY_NUC

Searched: 780561 seqs, 2137953050 residues

Database: GenEmbl.*

Word size: 0

Number of hits that pass the threshold : 1561122

1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: gb_v1.*
17: em_fun.*
18: em_htg.*
19: em_hum1.*
20: em_hum2.*
21: em_in.*
22: em_om.*
23: em_or.*
24: em_ov.*
25: em_pat.*
26: em_ph.*
27: em_pl.*
28: em_ro.*
29: em_sts.*
30: em_sy.*
31: em_un.*
32: em_v1.*
33: gb_htg1.*
34: gb_htg2.*
35: gb_in1.*
36: gb_in2.*
37: em_ba1.*
38: em_ba2.*
39: em_hum3.*
40: em_hum4.*
41: gb_pr4.*
42: gb_htg3.*
43: gb_htg4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|------------|---------------------|
| 1 | 2875.4 | 99.9 | 2877 | 41 | AF067864 | AF067864 Homo sapi |
| 2 | 711.8 | 24.7 | 227968 | 11 | AF053356 | AF053356 Homo sapi |
| 3 | 550.4 | 19.1 | 580 | 11 | HUMYR50H04 | AF085928 Homo sapi |
| 4 | 305.4 | 10.6 | 2500 | 4 | GCTFR | X55348 G.gallus mr |
| 5 | 280.8 | 9.8 | 5010 | 9 | HSTR | X01060 Human mRNA |
| 6 | 280.2 | 9.7 | 2826 | 10 | HDMTFR | M11507 Human trans |
| 7 | 256 | 8.9 | 2460 | 12 | CRUTRANREC | L19142 Cricetus |
| 8 | 247 | 8.6 | 3413 | 12 | RATFRFR | M58040 Rat transfe |
| 9 | 215 | 7.5 | 2292 | 12 | MMTRMRNA | X57349 M.musculus |
| 10 | 160.2 | 5.6 | 780 | 5 | I23809 | I23809 Sequence 27 |
| 11 | 146 | 5.1 | 250 | 13 | G14981 | G14981 human STS S |
| 12 | 127 | 4.4 | 540 | 5 | I23811 | I23811 Sequence 29 |
| 13 | 103.6 | 3.6 | 2570 | 12 | AF009921 | AF009921 Rattus no |
| 14 | 91.6 | 3.2 | 2320 | 10 | HSAN012371 | AJ012371 Homo sapi |
| 15 | 88.8 | 3.1 | 660 | 5 | I23810 | I23810 Sequence 28 |
| 16 | 73.4 | 2.6 | 3152 | 10 | HSAN012370 | AJ012370 Homo sapi |
| 17 | 67.8 | 2.4 | 1428 | 12 | AF039707 | AF039707 Rattus no |
| 18 | 67.8 | 2.4 | 2348 | 12 | AF040256 | AF040256 Rattus no |
| 19 | 67.8 | 2.4 | 2899 | 12 | RNU75973 | U75973 Rattus norv |
| 20 | 66.8 | 2.3 | 2603 | 12 | AF026380 | AF026380 Mus muscu |
| 21 | 57.8 | 2.0 | 1278 | 1 | ABIPDAGEN | X88853 A.brasiliens |
| 22 | 56.6 | 2.0 | 43632 | 1 | SC6A5 | AL049485 Streptomy |
| 23 | 56.2 | 2.0 | 2532 | 3 | AF050502 | AF050502 Sus scrof |
| 24 | 56 | 1.9 | 22449 | 1 | SPSNBCDE | Y11548 S.pistinae |
| 25 | 55.6 | 1.9 | 22449 | 1 | SPSNBCGEN | X98690 S.pistinae |
| 26 | 55.6 | 1.9 | 306 | 13 | G45215 | G45215 221292.1 ze |
| 27 | 54.8 | 1.9 | 189 | 12 | S81327 | S81327 transferlin |
| 28 | 53.8 | 1.9 | 2653 | 5 | I23794 | I23794 Sequence 1 |
| 29 | 53.8 | 1.9 | 2653 | 10 | HUMPSM | M99487 Human prost |
| 30 | 53.8 | 1.9 | 2518 | 41 | AF176574 | AF176574 Homo sapi |
| 31 | 53.6 | 1.9 | 4459 | 1 | TTPHST | X65609 T.thermophi |
| 32 | 53.6 | 1.9 | 7621 | 1 | TTPHSTPO | Y15464 Thermus the |
| 33 | 53.6 | 1.9 | 4171 | 1 | TTPHSTSC | Z12118 T.thermophi |
| 34 | 53 | 1.8 | 38995 | 1 | SCF34 | AL109974 Streptomy |
| 35 | 53 | 1.8 | 1189 | 9 | GCTRPPEX1 | Z11805 G.criassicu |
| 36 | 51.8 | 1.8 | 3508 | 2 | AF069761 | AF069761 Halobacte |
| 37 | 51.8 | 1.8 | 2014 | 2 | AzSPD | I26240 Azospirillum |
| 38 | 51.8 | 1.8 | 982 | 7 | ZMHSPI18K1 | X54075 Maize mRNA |
| 39 | 50.2 | 1.7 | 10295 | 1 | D63799 | D63799 Thermus the |
| 40 | 50.2 | 1.7 | 40745 | 1 | SC151 | AL108848 Streptomy |
| 41 | 49.8 | 1.7 | 24700 | 1 | SCB8 | AL035654 Streptomy |
| 42 | 48.8 | 1.7 | 40105 | 1 | SC5F2A | AL049587 Streptomy |
| 43 | 48.8 | 1.7 | 26440 | 1 | SC15 | AL049707 Streptomy |
| 44 | 48.6 | 1.7 | 843 | 7 | ZMHSPI18K2 | X54076 Maize mRNA |
| 45 | 48.2 | 1.7 | 4933 | 1 | ABIPDC | X99587 A.brasiliens |

ALIGNMENTS

| RESULT | 1 | ALIGNMENTS |
|------------|---------------------------------------------------------------------------------------------------|------------------------------|
| AF067864 | AF067864 | 2877 bp mRNA PRI 27-JUL-1999 |
| LOCUS | AF067864 | 2877 bp mRNA PRI 27-JUL-1999 |
| DEFINITION | Homo sapiens transferrin receptor 2 alpha (TFR2) mRNA, complete cds. | |
| ACCESSION | AF067864 | |
| NTD | 95596369 | |
| VERSION | AF067864.1 | GI:5596369 |
| KEYWORDS | | |
| SOURCE | human. | |
| ORGANISM | Homo sapiens | |
| REFERENCE | 1 (bases 1 to 2877) | |
| AUTHORS | Kawabata,H., Yang,R., Hirama,T., Vuong,P.T., Kawano,S., | |
| TITLE | Combart,A.F. and Koeffler,H.P. | |
| JOURNAL | Molecular cloning of transferrin receptor 2. A new member of the transferrin receptor-like family | |
| | J. Biol. Chem. 274 (30), 20826-20832 (1999) | |


```

Db 1501 GGGTACTGAGCGTGGCGACCTCAAGCGGTAGTGTAGCTGAGCCCTGGACACAGCGAGTG 1560
Qy 1561 ctggggagatgaagaatttcacgaagaagccccccttcgaagaatcattatgagat 1620
Db 1561 CTGGGGAGTATGACAAAGTTTCTATGCCAAGACAGCCCTTCTGACAAAGTCTCATGAGAGT 1620
Qy 1621 gtctctgaagagatgagatcttctcccaacaagatgggagagatctctctgaagaagtgatg 1680
Db 1621 GTCTCTGAAGAGATGAGATTTCTCCCAACACAGTGGGAGACTCTATGTAAGAGTGTGTG 1680
Qy 1681 ttcaacaatcccaagctggagatgctgaagtgatccgagccctacacatggagacaagtgcc 1740
Db 1681 TTCAACAATCCCAAGCTGGAGTGTGATGATGATCCGCGCCCTTACCAATGAGACAGAGTGC 1740
Qy 1741 taattcctcaagagccttggtggagagtcctcgctcgagttcttccttatttgagagacac 1800
Db 1741 TATTCTTCAAGCGCTTTGTGGAGTCCCTGCGTCAAGTTCTCTTATGAGAGACGAC 1800
Qy 1801 cagcctaccatctcctcgacacaaagagagacattatagaacctgataagtgctg 1860
Db 1801 CAGGCTTACCCATCTCTGACACAAAGAGAGACATTATGAAACCTGCAATAGGTGCTG 1860
Qy 1861 caagagccgctgcgcgcgcgtggcccaagcgctggccagctcgagggagagagctctctatc 1920
Db 1861 CAAGCGCGCTGCGCGCGCGCGCCGAGCCGCTGCGCCAGCTCGAGGAGCGCTCTCTATC 1920
Qy 1921 cggctcaagcagatcgctgctgccccctgacttcgcccctacagggagagagctgctc 1980
Db 1921 CGGCTCAGCCAGCATCGCTGCTGCGCCCTGACTTGGCGGTACGGGGAGCGTGCCTC 1980
Qy 1981 aggcacatcgagaaacctcaacagagttctcggggagactcaaggcccgaggctcgacctg 2040
Db 1981 AGGCACATCGGGAACCTCAACGAGTTCTGCGGAGACTCAAGGCCCGCGGCTGACCTG 2040
Qy 2041 cagtgggtgactcgagcgagggagagactacatccggcgagcgaagaagagcgagcagag 2100
Db 2041 CAGTGGGTGACTCGCGCGCGGGGAGACTACATCCGGCGGCGGAAGAGCTGCGGAGAG 2100
Qy 2101 atataagctcgagagagagagagacgagctgacacgcatgtataacgctgacataatg 2160
Db 2101 ATCTACAGCTCGGAGAGAGAGACGAGCTGACAGCATGTACACGTCGCGCATATG 2160
Qy 2161 cgggtggagtttacttcttcccttccagtaagtgctgcagacccgactcccgctccgacac 2220
Db 2161 CGGGTGAGTTTACTTCTTCCCACTAGCTGTCGCGCACCCGACTCCCGTTCGCGCAC 2220
Qy 2221 atcttcagcgctggagagacacacgctggcgccctgctggagacacactgagcgctgctg 2280
Db 2221 ATCTTCATGGCGCGTGGAGACACACGCTGGCGCCCTGCTGGACACCTGCGGCTGCTG 2280
Qy 2281 cgcctcaacagctcctggagaccccgaggagacactctctcactggtcttcagagagacgct 2340
Db 2281 CGCTCAACAGCTTCGGGAGACCCCGGGGACACCTCTCACTGGCTTCAGAGAGAGCGT 2340
Qy 2341 ttccggcgctagctagcctgctcactgagcgtgcaaggggagcgaatgagcttaagc 2400
Db 2341 TTCGGGCGTAGCTAGCGCTGCTCACTGAGCGTGCAGAGGGGAGCAATGCGCTTAGC 2400
Qy 2401 gggagatgctggagacattgataaactcttgagggccctgggagatctctacatcccgctc 2460
Db 2401 GGGGATGCTGGAACATTTGATACAACTTCTGAGGCGCTTGCGGATCTCTCAATCCCGCTC 2460
Qy 2461 ccccaagtcacaagacctcctcgcctcctgcttgatgatacagggctcagagagtgctca 2520
Db 2461 CCCAAGTCAAGAGCTCTGCTGCTCTGCTGATGATGATTCAGGGTCAAGGAGGTGCTCA 2520
Qy 2521 gattcacaactcattatgctatcaatttctcaataccctacacatctctccagagcc 2580
Db 2521 GATTCACACTCTATTCATCTGATCAATTTTCATTAACCCCTACACATCTCTCCAGGAGCC 2580
Qy 2581 cagacccagcagagatcatcacacaccagccctgcatgtgtagctgacccataatgta 2640

```

```

Db 2581 CAGACCCACAGACAGATATCCACACACACCCAGCCCTGTCATGTAGCTGACCCCTATATGTA 2640
Qy 2641 cggltacactgctcggttaatcagagagtagacatcccttcaatcacagcccttccctt 2700
Db 2641 CGGTCAATAGCTGCGTGAATACAGAGATAGATCATCCCTTAATACAGCCCTTCCCTTT 2700
Qy 2701 ctgggttccctccatctactagaagacacactcgtggaggttttctaaagcccttgagcctg 2760
Db 2701 CTGGGTCCTCTCATACTTACAGACACTCTGGAGGTTTCTTGGCCCTGGAGCTGGCC 2760
Qy 2761 agctctgtatgtggagagatcgcctgacacataagccttataggccaagtgatgctg 2820
Db 2761 AGCTCTGTATGTGGAGAGATCGCTGCGACACATAGCCTTATGGCCAAAGTGTGCTG 2820
Qy 2821 gtgaagggcggtggagtttcaatataataaaccactgatatcaataagccaanaa 2877
Db 2821 GTGAAGGGCGGTGGAGTTTCATATCAATCAATAAACCACTGATATCAATAAGCCAAA 2877

RESULT 2
AF053356 227968 bp DNA PRI 24-NOV-1998
LOCUS Homo sapiens chromosome 7q22 sequence, complete sequence.
DEFINITION AF053356
ACCESSION AF053356
NID G3135305
VERSION AF053356.1 GI:3135305
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 227968)
AUTHORS Gloeckner,G., Scherer,S., Schattevoy,R., Boright,A., Weber,J.,
Tsui,L.C. and Rosenthal,A.
TITLE Large-scale sequencing of two regions in human chromosome 7q22:
analysis of 650 kb of genomic sequence around the EPO and CMT1
loci reveals 17 genes
JOURNAL Genome Res. 8 (10), 1060-1073 (1998)
MEDLINE 99018118
REFERENCE 2 (bases 1 to 227968)
AUTHORS Gloeckner,G., Rosenthal,A., Drescher,B., Weber,J. and Schattevoy,R.
TITLE Direct Substitution
JOURNAL Submitted (05-MAY-1998) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q22"
/region="Region: 3763"
/complement(1..197)
/feature="SINE/Alu, pos., in, rep., 2..197"
/misc_feature 1..169722
/feature="PAC clone 37G3"
/feature="Region: 3763"
/complement(join<309..578,2314..2458,3620..3831,
14784..14928,15156..15357,16624..16804,18675..18769,
19928..20039,20325..20562,21131..21378,21722..21863,
22834..23036,24183..24361,24472..24582,26030..26173,
27852..27997,28543..28667,29738..29857,30045..30186,
32978..33158,35152..36676,37121..37205,37491..37762,
41053..41140,41322..41593,41773..41919,46462..46608,
46717..46795,47291..47382,48230..>48478)
/product="zonadhesin"
/complement(309..378)
/feature="100.000, excellent"
/complement(join<309..578,2314..2458,3620..3831,
14784..14928,15156..15357,16624..16804,18675..18769,
19928..20039,20325..20562,21131..21378,21722..21863,
22834..23036,24183..24361,24472..24582,26030..26173,

```

```

27852..27997,28543..28667,29738..29857,30045..30186,
32978..33158,35152..36676,37121..37205,37491..37762,
41053..41140,41322..41593,41773..41913,44662..46608,
46717..46795,47291..47382,48230..48383))
/codon_start=1
/product="zonadhesin"
/protein_id="AAC78790.1"
/db_xref="PID:g3135306"
/db_xref="GI:3135306"
/translacion="MLDLPKNARPGQKAVLLSPVSLSSGCLSFSEHYILRQSPGAAL
HIVASVLSIRKHTLFGSGDGPNNQAVSVNTAVGRIOFAVVGFGKPEPAVAVDT
SIAPGEGFPQCCEDENAHPCDWOTSGDGHVAGLKNKPNVHGAGGAFNPADVL
TQCFEDDAKPLCDMSQYASADEDMWRASGSPGSGHAGPGYNGEGSYHMSNPF
HRGSVARLSPDLMEQGPLCYHFAHMHGKLSWGQOLRLLLSGEGRRPDYLMKHMT
ORPSMMLTVTPAGFPLPLRLMEGEGSTAYVLDIALDALISIRGSCNRRGHLYLE
ADPSHAGQSVRLVSPRCAGDICVERAVAYMIGETMELLIGSPAGSPPLPK
ELVQSQRPMONTSVTVSGHQPMOLIFKRGQSNITASVMAFGLINPGCPKAVLP
RPPVSPVSTSGPSETTGLTENPTISTKKPTVSTIEKPTVEKPTVEKPTITEKP
TISTEKPTIESEKNMSEKPTIESEKPTIESEKPTIESEKPTIESEKPTISTEKPY
PTEPTPTTEETTYMEEPVPTPEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
EKPTIPTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
LTIPTERKLTIPTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
TSVETTISTEKLTIPMEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
IEETTISTEKLTIPTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
KPTISPEKLTIPTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
PSPATGIALVMSPHAPSTPMTSVIICTTTSSSGMSCPNARYESCQPSACKS
PRSPGCLRCREGCCVGNFELFSDNHCIOASSCCINDYIEPQAKMFTPCTHECR
WPSRVEQISQCGTHTVQLKNQOYGHYPAGATLVYDGPYVFDGHHFPMK
CTYILAPISKGVLFPRVATKNEQGEQVSCSKYVTLPESTVTLKLRRLVVG
QOVLPAIPSGVFLGASREVELQTEFGLRVDGQDLVTVSTVSGCLCGN
YDGSNDNDHLKDSGPAQDKEELGNSMOTDQDEQCKOVNPSQDSLSQSMG
PGFCGLVDTHTGPFETCLHYKASFPDSCMLDMCGROGLOHILCTHSTMTTCCOA
GHAYKPREPHEFCHELSSACPNNKSYSLCAKPCDPTCHSGFSCFSCRCVACECP
GFVLSGECIPRSQGLHPAGSYFKVGERMYKQCELCVCSNNRIRPCQWPCRCNP
EFCQODG6IYCGHAQGAATCA5GDPHVLTEDGALHMRGCTVTLRPCQWPCRCNP
FVVSATNENRGILLEVSYIKAVHTVDELISILRGCKVMALGRHALVPVLAAGRT
IRLSNNTLVLTNFGLOYRDXSHLVELTVPSYSGOLGCGEPGLRLKAGMT
RVGNYNNSLNDMLRPDRKLAGDSMOLGAAMKLPESSEPCFLVGGKPSQCSNSNA
DANKNCAILINPGQSPSQHQVYPPQSSPASCYHGGCGTGTGDTALCRSLQATASLC
AQQAQAPMARNRITFCRECPQSSYSPCSSPCPDTCS1NNRDCPKALPCASECECK
GHILSTSCVPLHCRASGVGCOLPGE5HYVSFDSCNHSIPACVLVYKCHPAMALP
NOICMALDGLHCRASGVGCOLPGE5HYVSFDSCNHSIPACVLVYKCHPAMALP
LEKISAKHEEGEGTEAFRLHEVYIDMVAOTVLOKGRVYLINSKQVTLISQIPGVS
SVKSSSYITVNIKIGVOYKRDGNHLLLEITPTVYKVGCMGCFNDEPEEDLIMS
DEKLANDSEFVNSMKDKDIDPSCOSLPVDQOQIPABQENPSCGRADIRAREKC
EALURAPVWQACASRIDLTPVLCANTLCFGLYALCALQALAFATQCSQSLKPP
LMRNSFCHCALSLPQTPPKPPPLCFPESLTLPLVPLECPAYSSYTNLPCSCPS
CGDLDRCEGAKVPSACAEICQPGYLSDEKCVPRQCCGKDAHGSIPLGKWS
SCDTEKCVCTGATGCGDFRCPSGHQOLTSDNSNSNCVSKSQC5YGGPRLTID
GFSYRLOGRMVYLIIKTYDVLPEGEVPLIVEGRNKMPPRSITLQEVITTVYGRKYQ
LQAGLELYVRAGPGP"
312..535
exon
/note="89.000, shadow-excellent"
/complement(327..435)
/misc_feature
/note="99, 1:xyz, 1..109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, length, =,
1090, P, =, 2.3e-37, S, =, 99"
/note="Region: Data base match"
/complement(327..435)
/misc_feature
/note="99, 173:xyz, 1..109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, length, =,
1090, P, =, 1.1e-32, S, =, 99"
/note="Region: Data base match"
/complement(327..435)
/misc_feature
/note="99, 1:xyz, 1..109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, length, =,
1090, P, =, 2.3e-37, S, =, 99"
/note="Region: Data base match"
/complement(328..435)
/misc_feature
/note="100, 3614:xyz, 1..106, of, GB|U83191|HSU83191_1,
Homo, sapiens, Eukaryotae, mitochondrial, eukaryotes,
Metazoa: Human, zonadhesin, (ZAN), mRNA, partial, cds.,
Sperm, membrane, protein., 4/97, length, =, 363, P, =,
27852..27997,28543..28667,29738..29857,30045..30186,
32978..33158,35152..36676,37121..37205,37491..37762,
41053..41140,41322..41593,41773..41913,44662..46608,
46717..46795,47291..47382,48230..48383))
/codon_start=1
/product="zonadhesin"
/protein_id="AAC78790.1"
/db_xref="PID:g3135306"
/db_xref="GI:3135306"
/translacion="MLDLPKNARPGQKAVLLSPVSLSSGCLSFSEHYILRQSPGAAL
HIVASVLSIRKHTLFGSGDGPNNQAVSVNTAVGRIOFAVVGFGKPEPAVAVDT
SIAPGEGFPQCCEDENAHPCDWOTSGDGHVAGLKNKPNVHGAGGAFNPADVL
TQCFEDDAKPLCDMSQYASADEDMWRASGSPGSGHAGPGYNGEGSYHMSNPF
HRGSVARLSPDLMEQGPLCYHFAHMHGKLSWGQOLRLLLSGEGRRPDYLMKHMT
ORPSMMLTVTPAGFPLPLRLMEGEGSTAYVLDIALDALISIRGSCNRRGHLYLE
ADPSHAGQSVRLVSPRCAGDICVERAVAYMIGETMELLIGSPAGSPPLPK
ELVQSQRPMONTSVTVSGHQPMOLIFKRGQSNITASVMAFGLINPGCPKAVLP
RPPVSPVSTSGPSETTGLTENPTISTKKPTVSTIEKPTVEKPTVEKPTITEKP
TISTEKPTIESEKNMSEKPTIESEKPTIESEKPTIESEKPTIESEKPTISTEKPY
PTEPTPTTEETTYMEEPVPTPEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
EKPTIPTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
LTIPTERKLTIPTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
TSVETTISTEKLTIPMEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
IEETTISTEKLTIPTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
KPTISPEKLTIPTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
PSPATGIALVMSPHAPSTPMTSVIICTTTSSSGMSCPNARYESCQPSACKS
PRSPGCLRCREGCCVGNFELFSDNHCIOASSCCINDYIEPQAKMFTPCTHECR
WPSRVEQISQCGTHTVQLKNQOYGHYPAGATLVYDGPYVFDGHHFPMK
CTYILAPISKGVLFPRVATKNEQGEQVSCSKYVTLPESTVTLKLRRLVVG
QOVLPAIPSGVFLGASREVELQTEFGLRVDGQDLVTVSTVSGCLCGN
YDGSNDNDHLKDSGPAQDKEELGNSMOTDQDEQCKOVNPSQDSLSQSMG
PGFCGLVDTHTGPFETCLHYKASFPDSCMLDMCGROGLOHILCTHSTMTTCCOA
GHAYKPREPHEFCHELSSACPNNKSYSLCAKPCDPTCHSGFSCFSCRCVACECP
GFVLSGECIPRSQGLHPAGSYFKVGERMYKQCELCVCSNNRIRPCQWPCRCNP
EFCQODG6IYCGHAQGAATCA5GDPHVLTEDGALHMRGCTVTLRPCQWPCRCNP
FVVSATNENRGILLEVSYIKAVHTVDELISILRGCKVMALGRHALVPVLAAGRT
IRLSNNTLVLTNFGLOYRDXSHLVELTVPSYSGOLGCGEPGLRLKAGMT
RVGNYNNSLNDMLRPDRKLAGDSMOLGAAMKLPESSEPCFLVGGKPSQCSNSNA
DANKNCAILINPGQSPSQHQVYPPQSSPASCYHGGCGTGTGDTALCRSLQATASLC
AQQAQAPMARNRITFCRECPQSSYSPCSSPCPDTCS1NNRDCPKALPCASECECK
GHILSTSCVPLHCRASGVGCOLPGE5HYVSFDSCNHSIPACVLVYKCHPAMALP
NOICMALDGLHCRASGVGCOLPGE5HYVSFDSCNHSIPACVLVYKCHPAMALP
LEKISAKHEEGEGTEAFRLHEVYIDMVAOTVLOKGRVYLINSKQVTLISQIPGVS
SVKSSSYITVNIKIGVOYKRDGNHLLLEITPTVYKVGCMGCFNDEPEEDLIMS
DEKLANDSEFVNSMKDKDIDPSCOSLPVDQOQIPABQENPSCGRADIRAREKC
EALURAPVWQACASRIDLTPVLCANTLCFGLYALCALQALAFATQCSQSLKPP
LMRNSFCHCALSLPQTPPKPPPLCFPESLTLPLVPLECPAYSSYTNLPCSCPS
CGDLDRCEGAKVPSACAEICQPGYLSDEKCVPRQCCGKDAHGSIPLGKWS
SCDTEKCVCTGATGCGDFRCPSGHQOLTSDNSNSNCVSKSQC5YGGPRLTID
GFSYRLOGRMVYLIIKTYDVLPEGEVPLIVEGRNKMPPRSITLQEVITTVYGRKYQ
LQAGLELYVRAGPGP"
312..535
exon
/note="89.000, shadow-excellent"
/complement(327..435)
/misc_feature
/note="99, 1:xyz, 1..109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, length, =,
1090, P, =, 2.3e-37, S, =, 99"
/note="Region: Data base match"
/complement(327..435)
/misc_feature
/note="99, 173:xyz, 1..109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, length, =,
1090, P, =, 1.1e-32, S, =, 99"
/note="Region: Data base match"
/complement(327..435)
/misc_feature
/note="99, 1:xyz, 1..109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, length, =,
1090, P, =, 2.3e-37, S, =, 99"
/note="Region: Data base match"
/complement(328..435)
/misc_feature
/note="100, 3614:xyz, 1..106, of, GB|U83191|HSU83191_1,
Homo, sapiens, Eukaryotae, mitochondrial, eukaryotes,
Metazoa: Human, zonadhesin, (ZAN), mRNA, partial, cds.,
Sperm, membrane, protein., 4/97, length, =, 363, P, =,
1.9e-14, S, =, 100"
/note="Region: Data base match"
/complement(328..576)
/misc_feature
/note="81, 6:xyz, 5899..6145, of, GB|U40024|SSU40024_1,
Sus, scrofa, Eukaryotae, mitochondrial, eukaryotes,
Metazoa: Sus, scrofa, zonadhesin, (Zan), mRNA, complete,
cds., The, amino, acid, sequence, predicts, seven,
extracellular, domains., Including, a, mucin-like, domain,
and, five, domains, homologous, to, the, D-do., ' '
11/95, length, =, 2476, P, =, 0.0e+00, S, =, 81"
/note="Region: Data base match"
/complement(331..578)
exon
/note="24.00, 82aa,
SEQSVTGDPRVLTFFDSFSTRLOGRMVYLIIKTYDVLPEGEVPLIVEGRNKMPPRSS
IFLQEVITTVYGRKYQLQAGLELY"
/complement(331..578)
/misc_feature
/note="1.000, 0.62, 0.44, 0.40, 1.22, 0.54, 0.57, 0.60"
/complement(339..588)
promoter
/note="7.43, TATA:, 367, CAP:, 337"
/complement(345..561)
exon
/note="98, prediction:, 0%, 98%"
/misc_feature
/note="80, 3119:xyz, 1..166, of, GB|U83190|MMU83190_1,
Mus, musculus, Eukaryotae, mitochondrial, eukaryotes,
Metazoa: Mus, musculus, zonadhesin, (zan), mRNA,
partial, cds., Sperm, membrane, protein., 4/97, length, =,
513, P, =, 2.0e-24, S, =, 80"
/note="Region: Data base match"
/complement(475..567)
/misc_feature
/note="70, 3642:xyz, 856..946, of, GB|U941321|HUMIMUCB_1,
Homo, sapiens, Eukaryotae, mitochondrial, eukaryotes,
Metazoa: Human, mucin, 2, (MUC2), mRNA, sequence., 1/95,
length, =, 984, P, =, 3.2e-14, S, =, 70"
/note="Region: Data base match"
/complement(475..567)
/misc_feature
/note="70, 1807:xyz, 13441..13531, of,
GB|U121998|HMMUC2X_1, Homo, sapiens, Mammalia:, Homo,
sapiens, Intestinal, mucin, (MUC2), mRNA, complete, cds.,
1/98, length, =, 5179, P, =, 1.5e-30, S, =, 70"
/note="Region: Data base match"
606..903
repeat_region
/note="SINE/Alu, pos., in, rep.: 1..296"
/rpt_family="Alusx"
1004..1039
repeat_region
/note="83.30, %, 22, length:, 2, #: 18, seq:, tt"
/rpt_type=tandem
/rpt_unit=score:
/complement(1021..1133)
repeat_region
/note="SINE/Alu, pos., in, rep.: 3..300"
/rpt_family="Aluuo"
/complement(1374..1684)
repeat_region
/note="SINE/Alu, pos., in, rep.: 1..302"
/rpt_family="Aluuo"
1745..2050
repeat_region
/note="SINE/Alu, pos., in, rep.: 1..302"
/rpt_family="Aluuo"
2026..2275
promoter
/note="7.13, TATA:, -1, CAP:, -1"
2166..2519
misc_feature
/note="14.60, 354, bp"
promoter
/note="5.62, TATA:, -1, CAP:, -1"
/complement(2314..2458)
exon
/note="0.850, 0.45, 0.46, 0.56, 1.11, 0.49, 0.54, 0.53"
/complement(2314..2458)
misc_feature
/note="94.000, excellent"
/complement(2354..2464)
/misc_feature
/note="75, 26:xyz, 5749..5857, of, GB|U40024|SSU40024_1,
Sus, scrofa, Eukaryotae, mitochondrial, eukaryotes,
Metazoa: Sus, scrofa, zonadhesin, (Zan), mRNA, complete,
cds., The, amino, acid, sequence, predicts, seven,

```

extracellular, domains, including, a, mucin-like, domain, and, five, domains, homologous, to, the, D-do., . . .

Query Match 24.7%; Score 711.8; DB 11: Length 227968;

Best Local Similarity 99.7%; Pred. No. 1.2e-106;

Matches 713; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2160 gggggtgaggttacttcttccagtagcgtgtgcccagccgctcccggtccgcca 2219
 |||||||
 Db 154517 GcAGGtGgAGtTCTACTCTTCTCCAGTAGcGTGcCCAGcGcAGCTCCCGTTCGcCA 154576
 |||||||
 Qy 2220 catcttcattgagccgtggaagaccacacagcgtgggcccctgtgagcaacctgtggcctg 2279
 |||||||
 Db 154577 CATCTTCATGAGGCGCGTGGAGACACACAGCGTGGCGCCCTGTGACCACTGCGGCTGCT 154636
 |||||||
 Qy 2280 ggcgtccaaagagctccggagaccccgggggcaccctccctcactggtctccagaagacg 2339
 |||||||
 Db 154637 GCGCTCCAAcAGCTCCGGGACCCCGGGGGCAGCTCTCCACTGCTTCCAGAGAGCCG 154696
 |||||||
 Qy 2340 ttccggggtcagctagccctgtctcactgtgagcgtgcaagggcgagccaatgctgtag 2399
 |||||||
 Db 154697 TTTCCGGCGTCACTAGCCCTGCTCACCTGAGCGTGCAGAGGGGCAAGCCAACTCGCTTAG 154756
 |||||||
 Qy 2400 cgggagatgtctggaacattgataacaacctctgagggccctggggatccctacatcccgct 2459
 |||||||
 Db 154757 CCGGGATGTCTGGAACATTATTAACAACCTTCTAGAGCCCTGGGGATCTCAATCCCGGT 154816
 |||||||
 Qy 2460 ccccgagtcgaagctcctcctctgtcctcgtctgtaattgattcagaggtcagaggtgctc 2519
 |||||||
 Db 154817 CCCCAGTCAAGAGAGCTCTCTGCTCCCTCGCTGAATGATTCAGGGTCAAGGAGGTGCTC 154876
 |||||||
 Qy 2520 agagtcacactctcattgtcgtatcaatcttctacacctacacatctctccagggagc 2579
 |||||||
 Db 154877 AGAGTCCACCTCTCATTTGCTGATCAATTTCTCATTTACCTTCACACATCTCCACGAGC 154936
 |||||||
 Qy 2580 ccgagaccccgacagatatccacacacaccccgagcccgaggttaggtgacctaattgt 2639
 |||||||
 Db 154937 CCGAGACCCCGAGCAGATATCCACACACCCAGCCCTGCGAGTGTAGCTGACCTTAATGTG 154996
 |||||||
 Qy 2640 acggcatactgtcgtgttaatacagagtagcatccctcaatacagcccttccctc 2699
 |||||||
 Db 154997 ACGGTATACTGTGCGTTAATCAGAGTAGATCCCTTCAATCAGCCGCTTCCCTT 155056
 |||||||
 Qy 2700 tctgggggtcctcattactatcagagcaacctctggaggttctgtaagccctggagcctg 2759
 |||||||
 Db 155057 TCTGGGGTCTCTCATCTAGAGACCACTCTGGAGGTTGTGTAAGCCCTGAGCACTGCG 155116
 |||||||
 Qy 2760 cagctctgttagtggagagatcgtctggcacccatagcctatggcaaacaggtgtgt 2819
 |||||||
 Db 155117 CAGCTCTGTTAGTGGAGAGATGCGTGGCACATAGCCTTATGGCCAAAGGTGTGTGT 155176
 |||||||
 Qy 2820 ggtgaagagggcgtggaggttcaataataaaccacccgtatataaagca 2874
 |||||||
 Db 155177 GGTGAAGGGGCGTGGAGTTCAATATCAATTAACCAACCGATATCATTAAGCCA 155231
 |||||||

Schurk, R., Ritter, E., Kohn, S., Swaller, T., Behymer, K., Hillier, L., Wilson, R. and Waterston, R. Full Clone Sequencing of the Longest Available Member from Each Unigene Cluster Unpublished 2 (bases 1 to 580) Waterston, R. Submitted (24-Aug-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA SUBMITTED BY: Genome Sequencing Center Department of Genetics Washington University St. Louis MO 63108, USA http://genome.wustl.edu/gsc mailto:estewatson.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived. We have tried to obtain double-stranded, or double chemistry sequence across the entire clone, but potentially, there are areas in the sequence where this level of coverage was not achieved. Nevertheless, we are confident of the accuracy of this sequence as all regions of low quality, as defined by PHRAP (P. Green, in preparation), were visually inspected and edited accordingly. The consensus quality values for this sequence have been submitted separately.

SIMILARITY INFORMATION:
 probable match to Homo sapiens protein AF053356 (PID:g3135312) transferrin-receptor2

FEATURES
 source location/Qualifiers
 1..580 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Soares_fetal_liver_spleen_inFLS"
 /clone="IMAGE:208759"
 1..510 /note="probable match to Homo sapiens protein AF053356 (PID:g3135312)"
 43..549 /note="similar to Homo sapiens protein P02786 (PID:g136378)"
 43..549 /note="similar to Rattus norvegicus protein A34549 (PID:g112425)"
 43..549 /note="similar to Mus musculus protein S29548 (PID:g346962)"
 43..549 /note="similar to Cricetus griseus protein A48592 (PID:g539790)"

misc_feature
 misc_feature
 misc_feature
 misc_feature

BASE COUNT 120 a 198 c 168 g 94 t

ORIGIN

Qy 544 gggcctcagcagacattgcggcgccctccgcccagagctggagcagcgtgtg 603
 |||||||
 Db 1 GCGGCTCTGACTAGAGACATTCGCGGCGCTTCCCGCAGAGGTGACACGtGtG 60
 |||||||
 Qy 604 accgacagcactacgttggggctgcaattcccgatccggtacacccacacccctggac 663
 |||||||
 Db 61 ACGGACACGACAGTACGTGGGGCGTCAATTCGCGGATCCGCTCACCCCAACACCTGAC 120
 |||||||
 Qy 664 tgggtcagatgagggcggaaggtcggagagcagctgcgctgtgaggaacccctgacgtctac 723
 |||||||

Query Match 19.1%; Score 550.4; DB 11: Length 580;
 Best Local Similarity 98.9%; Pred. No. 4.4e-80;
 Matches 554; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

|||||
Db 121 TGGGTCGATGAGCGCGGAGAGTCCGAGAGCACTGCCCTGAGAGACCTTACGTCTTC 180
Oy 724 tgcctctacagcgcacatgcggcaacgtcacaggagagctgtgtacgcccactacaggag 783
Db 181 TGCCCCCTACAGCGCATCGGCAACGTCAACGGAGAGCTGTGTACGCCACTACGGGCGG 240
Oy 784 ccggaagacctcagagactctgaggccaggagcgtgtgacctgaggccgctgtctgctg 843
Db 241 CCCGAAGACCTCAGAGACCTGCGGCGCAGGGGCGTGAGATCCAGTGGCGGCTGCTGCTG 300
Oy 844 gtgcgcgtgtggtgtatcagctctgcgcaggaagtgtacccaatgtcgaagactctggggt 903
Db 301 GCGCGCGTGGGGGTATACGCTTCGCCCGCAGAGGTGACCAATGCTCAGACTTCGGGGCT 360
Oy 904 caaggagtgtctatataccacagaccagcgaactctccacagaccaccacccaagccaagc 963
Db 361 CAGGAGAGTCTCATATACCCAGAGCGGAGGACTTCTCCACGAGACCCACCAAGCCAAAGC 420
Oy 964 ctgtccagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1023
Db 421 CTGTCCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 480
Oy 1024 cctgtctccctctctctcaatacaaacccagctccctccagctgtgcatcagccttccc 1083
Db 481 CCTGGCTTCCCTCTCTCTCAATCAAAACCAAGTTCCTCCAGTTCATCATCAGGCTTCCC 540
Oy 1084 agcatccagccagccagccat 1103
Db 541 AGCATCCACGCGCCGAGCCT 560

RESULT 4

LOCUS GGTFR 2500 bp mRNA VRT 30-SEP-1991
DEFINITION G.gallus mRNA for transferrin receptor.
ACCESSION X55348
VERSION 963806
KEYWORDS X55348.1 GI:63806
transferrin receptor.
SOURCE chicken.
ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 2500)

REFERENCE
1 Chan, L.N.L.
Direct Submission
Submitted (01-NOV-1990) Chan L.N.L., University of Texas Medical
Branch, Department of Human Biological Chemistry and Genetics,
Univ. Texas Medical Br., Galveston, TX 77550 U.S.A

REFERENCE
2 (bases 1 to 2500)

REFERENCE
Gerhardt, E.M. and Chan, L.N.L.
Structure and Organization of the Chicken Transferrin Receptor and
its cDNA Sequence
Unpublished

JOURNAL
FEATURES
source

Location/Qualifiers
1..2500
/organism="Gallus gallus"
/db_xref="taxon:9031"
/cell_type="erythroblast"
/clone_lib="gt11.chicken erythroblast"
/clone="CTRBV4"
79..2415
/standard_name="CTRCDNA1"
/codon_start=1
/product="chicken transferrin receptor"
/protein_id="CA39035.1"
/db_xref="PID:963807"
/db_xref="GI:63807"
/db_xref="SPTREMBL:Q90997"
/translation="MDHARALSNLFVSPNSYTRFSJAROTDGNHVENKLSADDE
EGDIERPEHHVMQPORNGKRLCFVIAVALLLILGFLIGLSTYGRTEILARCO
DGSGETPTPASTYLVDEGTVEEIIQGPVTFWPELKAMLSKLSKNTLVNTRWRY

GVDSFEAGAEPTMTATYIEEFNRNPLDKYWNDEHYIKLIVRGSTKNQVSIKNGKE
ILTEPDALIVYSESGYSGRPVYNTGLKADFEIQLKAVASLNGITIVRAGKITLLE
KVANAKFAGAGVAVLMDVSLMYGTDILIFGNAHLGTGDPYFPFSPHNTOPVE
SSGLPHIAYGTIISSSAARLFKSKMDGTCEBGMKGAHSCVYTKQSOIMVKIDVN
SMKDRKILNIFGAIOGFEEDRYAVIQAOSDQPGVAKAGTGAILELAVLSIDY
KNBGRKRRIIFASMSAGDYGVANGATEMEGYSAMHAAFTYISDAVPLGSHVK
ISAPILYIMSELYIMKGRNPAVSESLYRIKPDWKAIVPLDGNAAAPFLAYSGI
PVLSFGYRNKDEEERFLDTKGTITENLRKIDNDALAAVAAGVAAALRTHDEL
LDIGRSSEILATVOEFLPYIKREIREGLTLDWLFARGFQRAVTLRDIANSDE
NRYLRALNDNRMMKVEYDFLSPYISPADVEFRHIFKRGHTLRSLEHDLTKTNS
SYDNLRLREOLALATWTIKRPMPEWEVYSKLTIMNSRH"

BASE COUNT 650 a 545 c 707 g 598 t

ORIGIN

Query Match 10.6%; Score 305.4; DB 4; Length 2500;
Best Local Similarity 52.2%; Pred. No. 1.2e-40;
Matches 833; Conservative 0; Mismatches 716; Indels 48; Gaps 5;

Oy 704 tggaggaacctgaagctactctgcccctacagcgcacatgagcaagctcagggagagctg 763
Db 737 TGGAGACTCTCGATGCTATCGTTGCATACACGAGAGTGCTCTGTTCTGCAAAACCTG 796

Oy 764 tgtacgcccactacggggcgccggaagacctgcagagactctgcggccagggcgctgagtc 823
Db 797 TCTATGTAACTACGCGCTGAAAAAAGATTGTGAGATATACAGAGGCTCGTGCCTTAC 856

Oy 824 cagtgccgcctctgtctgtgtgcgcgtgtgggtgtatcagcttcgcgcccaaggttgacca 883
Db 857 TGAATGAAACCATAGTACTTGTCTAGAGCTGGAAAAATTAACACTTGTGAGAAAGTTGCA 916

Oy 884 atgtcagagactctggggtcagaagatgtcctatataccacagaccagcgagactctccc 943
Db 917 ATGCCAAAGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 957

Oy 944 aggaaccaccacaaagcctgtccagcagcagcagcagcagcagcagcagcagcagcagc 1003
Db 957 -GGATTCTATGTATGGAATTAACAGATACACTTATCCATTGGAGATCCCACTTG 1015

Oy 1004 gaactggaacccctacacacactgtctccctccctcaatacaaacccagctccctcag 1063
Db 1016 GAACGTGAGACCTTACACCCAGGCTTCCTGTGTTCAACACACACCCAGTTCCACCAG 1075

Oy 1064 ttgcatatcagcctctccagcatccagccagccacatcagtgcaagatctgtctccc 1123
Db 1076 TTGAATCTTCAGGACTACCCCACTTGTCTTCAGAGACATCTTCAGCATCAGCATCAG 1135

Oy 1124 gctctctgaggaagctcaaaagccctgtgcccaccagaaatgtcagggagcctctag 1183
Db 1136 GCGTGTTCAGCAAAATGAGATGAGACACATCTCTGAAAGTTGAAAGTCCGATCCATT 1195

Oy 1184 gct-----ccctataccctgtggcccgccagcagctgcagctagtgtcaacatc 1237
Db 1196 CCTGTAAAGTGACAAACAAAGAGGAGGACCGAATATGTAACATAGATGTAACAATT 1255

Oy 1238 acaggaacctccaccccatcaacaacatctcgtgtcatcgaagggcgctcagagccag 1297
Db 1256 CCATGAAAGACAGAGAAAGATTGTGAACATCTTCGGTGTATCCAGGAGATTGAAACACTG 1315

Oy 1298 atcactagctgtatctcggggccagagagatgatcagggggccagggagcagctaatccg 1357
Db 1316 ATCGGTATGTGTGATTGAGCCCAAGAGACTCTCGGGGCCAGGAGTGGCTTAAGCTG 1375

Oy 1358 ctgtgggagcgcctatactcctgtgagctgtgctgagcacttctccatcagtgtagaac- 1417
Db 1376 GCACGTGAAGATGCTATATTTGGAACCTTCCCGTGTATCTCAGACATATGTAATAAA 1435

Oy 1417 --ggttcggcccgccaggaagctctctcatcagctgaggaaggggtgactcttgaa 1474
Db 1436 AGGGCTACAAACCGAGGGGAGACATCACTTGTAGCTGAGATGCGAGAACTACGAGAG 1495

Oy 1475 gcttgagctcagcagagtgtagaagctacccagctgtctgcaactcaaacgcgtag 1534
|||||

| QY | 1496 | CTGAGGCTGCTACGATGCGTGGAGGGGCTACTGCTGCATCTCATGTCACCAACCTTCA | 1555 |
|----|------|--------------------------------------------------------------|------|
| QY | 1535 | tgtagctgagccctgagacaagcagtgctgggagtagacaagttcatgccaagaccagcc | 1594 |
| Db | 1556 | CTTACATCAGCTTGGATGCTCCACTCTGGGAGACCAACATGATTTCTCCAGCC | 1615 |
| QY | 1595 | ccctctgacaagctcatatgagagtgctcctgaagcaggtgagatctcccaaccagtg | 1654 |
| Db | 1616 | CTTGTGCTGATATGCTCTGGGAGATATTAAGAGGGGGTGAAMAATCCAGACAGCTCT | 1675 |
| QY | 1655 | gagcagactctcatgacaggtggtgttcaaccaatcccaagctggagatgagtgatcc | 1714 |
| Db | 1676 | CAGAGACCTCTATACAGACTTGGCCAGACTGGGTAAGAGCTGTGT----- | 1726 |
| QY | 1715 | gagccctaccacatgagacagcagtgctcctcctcaagcctctgtggagatctccg | 1774 |
| Db | 1726 | -CCTTTGGCTGATATGCTCTGGGAGATATTAAGAGGGGGTGAAMAATCCAGACAGCTCT | 1783 |
| QY | 1775 | tcgagctcctcatatgagagcagcagcctcctcctcctcctcctcctcctcctcctc | 1834 |
| Db | 1784 | TGCTTTGCTTCTTACATTAAGATAGAGATATGCTCTGGGAGACTTAAGAGGTGACA | 1843 |
| QY | 1835 | cttatgacaacctgcatagagtgctgcaagcgcgcctgcccgcgtgagccagcgcgtg | 1894 |
| Db | 1844 | CACGTGAGAACTGAGGAAATATGATAT-----CTGAGATCTCTTGGCTCTGCTG | 1897 |
| QY | 1895 | cccagctcgaagcagcctcctcctcctcctcctcctcctcctcctcctcctcctcct | 1954 |
| Db | 1898 | CAGAGAGCTGAGCAAGCAGCTCTCAGCGTACAGCCATGATGAGCTCTCTGAGACA | 1957 |
| QY | 1955 | tcggcgcctcgggagagctgctcctcagcatcgggaacctcaagagttcctctggg | 2014 |
| Db | 1958 | TGGGAGATACAGGAAATTAAGTATGATGCAAGAGAGGATTTTCTTACATTAAG | 2017 |
| QY | 2015 | acctcaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc | 2074 |
| Db | 2018 | AAGTGGGAGAGCTGGGCTTGAAGCTTGAAGCTTGTCTTCTTCTTCTTCTTCTTCTT | 2077 |
| QY | 2075 | gggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc | 2134 |
| Db | 2078 | GAGCTGTAAGTGCACCTGAGAAAGACATGCAACAGTGAAGGGGAGAACAGGCTATCC | 2137 |
| QY | 2135 | caagcatgctcaagctgcatatgctgaggtgagtgctcctcctcctcctcctcctcct | 2194 |
| Db | 2138 | GCAAGGCCCTGAATGACAGATGATGAGGATGATGATGATGATGATGATGATGATGAT | 2197 |
| QY | 2195 | cgccagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc | 2254 |
| Db | 2198 | CACCAAAAGATGTCCTTTTCGCCACATCTTTTGGGCAAAAGGCCCCACACCTGCGGA | 2257 |
| QY | 2255 | ccctctgagcagcagctgctgctgctgctgctgctgctgctgctgctgctgctgctg | 2291 |
| Db | 2258 | GTCGTGGAGAGATCTGACAGCTGTTGAAGACACAGAG 2294 | |

RESULT 5
 HSTR 5010 bp mRNA PRI 11-APR-1995
 LOCUS Human mRNA for transferrin receptor.
 DEFINITION X01060
 ACCESSION X01060
 NID 337432
 VERSION X01060.1 GI:37432
 KEYWORDS transferrin receptor.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 5010)
 AUTHORS Schneider,C., Owen,M.J., Banville,D. and Williams,J.G.
 TITLE Primary structure of human transferrin receptor deduced from the
 JOURNAL Nature 311 (5987), 675-678 (1984)
 MEDLINE 85012743

| QY | 839 | tgctgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg | 898 |
|----|------|--------------------------------------------------------------|------|
| Db | 1027 | TGATTGTCAAGCAGCAAGAAATACCTTTCAAGAAAGTTCGAAAGATGCTTA | 1086 |
| QY | 899 | gggtcagagagtgctatataccagagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc | 958 |
| Db | 1087 | ATGCAATTTGTTGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT | 1128 |
| QY | 959 | caagcctgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc | 1018 |
| Db | 1129 | CAATTGTATACGAGAACTTCAATCTTTTGGAGATGCTATCTGGGAGACAGTACCTT | 1188 |
| QY | 1019 | acacacctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct | 1078 |
| Db | 1189 | ACACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT | 1248 |
| QY | 1079 | ttccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc | 1138 |
| Db | 1249 | TGCTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT | 1308 |
| QY | 1139 | tcaaaag-ccctgtgcccacaaagatgagcagcagcagcagcagcagcagcagcagcag | 1197 |
| Db | 1309 | TGGAAGAGAGCTGCTCCTTGATGATGATGATGATGATGATGATGATGATGATGATGAT | 1368 |
| QY | 1198 | ctgggcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc | 1257 |
| Db | 1369 | AAAGCAAGATG-----TGAAGCTCAGCTGATGATGATGATGATGATGATGATGATGAT | 1421 |
| QY | 1258 | aaacaacatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg | 1317 |
| Db | 1422 | CTTAAATCTTTGATTAAGGCTTTGTAGAACCAATGATGATGATGATGATGATGATGAT | 1481 |

COMMENT Data kindly reviewed (19-FEB-1986) by C. Schneider.
 FEATURES
 Location/Qualifiers
 1..5010
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 CDS
 264..2546
 /note="put. transferrin receptor (aa 1-760)"
 /codon_start=1
 /protein_id="CAA25527.1"
 /db_xref="pid:g37433"
 /db_xref="gi:37433"
 /db_xref="SWISS-PROT:P02786"
 /translation="MMDQARSAPSNLFGGEPISYTRFSLAROVGDNSHVENKLAVIDE
 ENADNTRANVTKPKRCSSICQYTAIVFLLIGFMIGLYCQGVPEKTECELA
 GTESPVREPEDEFPARARLYMDLKRKSEKIDSDFTSTIKLHNSVYVREASQ
 KENALVYENQREFELSKVNRDQHVYKIOYKDSQNSVVIYDKRGLVYLVENQSG
 YVAASKAATVTKLVANNGTKRDEFDTLPNGSIVIRAGITFPAEKVANSINA
 IGVLYMDQTKPIYNAELSFCHALITGDDPYTTPFPFNHQPFPSSSGLPNIPV
 OTISRAAEKLFQNNBDDPSDKTIDSTQRMVTSBKVKLVISNVLKIKILNITGV
 IKGVEPDHYVAVAGQADWAGBGAAGVGTALLKLAQFSDMVLDKGFQPSRIIF
 ASMSADGFSVGAATEMLEGYSLSHLKATYINLDKAVIGTSPFKVASPLLYTLIEK
 TMQNVKHPVYGOFLYODSNMWSKVEKTLIDNAAPFLAYSGIPAVSFCEPDYPL
 GTMDTYKELIRIPELNVARAFAVAGQFVYKILHVDLNDYRYSNLSOLSEFVRD
 LNOYRADIREKLSIOMLYSARGDFPRAISRLITDQNAEKTDREPMKLNDRMAYVE
 YHPLSPYVPSKSPFRHVFWSGSHLPLALLEMLKRNKONGAFNETLIRNOLATMT
 TIQGANALSGDWDIDNEF"
 BASE COUNT 1399 a 977 c 1155 g 1479 t
 ORIGIN
 Query Match 9.8%; Score 280.8; DB 9; Length 5010;
 Best Local Similarity 54.1%; Pred. No. 9.4e-37;
 Matches 781; Conservative 0; Mismatches 607; Indels 56; Gaps 8;

```

OY 1318 gcccaaggatgcatgagggcccaaggagcagcgaatcccgctggtggagcgctatacc 1377
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1482 gcccaaggatgcatgagggcccttgagctgcaaaatccgggtatgacacagctctccta 1541
OY 1378 ctgagagctggtcgagaccccttctcccaatg--tgaagaacggctccggcccgacaga 1434
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1542 tggaaacttgcccaagatggtctcagatgctttaaagaaggggttcagccacacaca 1601
OY 1435 agctcccttccatcagctcgaggaacggtgtgtgactttggaagcggtggctccacagagtg 1494
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1602 accattatcttggccaggttgagtgctgagactttggatcggttgctccactgatatg 1661
OY 1495 ctgaagagctaccctcagctgctgacccctcaagcgtagtgaagtgagcctggagcaac 1554
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1662 ctgaagagctacccttgcgtccctgcctttaaaggctttacttatattatctggatpaa 1721
OY 1555 gcaagtgtggtggatgacagattcctcagcaagaccagccctctcagcaagctcatt 1614
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1722 gcggttcttggtacacgcaacttcaaggttttgcacagccactgtgtatagcgttat 1781
OY 1615 gagatgtctctgaagcaggtgatctcccaaccagatgggcagactctatagaacg 1674
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1782 gagaaaaaatgcamaaatgtgaagcattcc--ggttactggcgaatttctatcagaga- 1838
OY 1675 gtggtgtcccaatcccaatccgagtgatgagtgatgctcgccctccatccatgagcagc 1734
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1838 -----cagcaactggccacgaaagttagaactctcttagacatt 1880
OY 1735 agtgcattctcctcagcgcttctgagagtcctcgctcgagctccttctatgag 1794
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1881 gctgttcttcccttctcctcattatctggaatcccgagcttcttctgttttcgag 1940
OY 1795 gacgcacagcctaccatctctctgcacaaagaagagaaacttatagaactgtcaag 1854
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1941 ---gacacagattatcttatttgggtaccacacacactatgaagacattgag 1997
OY 1855 gtgcgcagagccgcctgcgcgcgtgagcccaagcggtgagccagctcagagcgagc 1914
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1998 aggatttccg---agttgaacaaagtgccacagcagctgacagagcttgcgtgctgattc 2054
OY 1915 ctcatccgcctcagccacgacatgcctgctgcctccctcgaactcgccgctacggggagcgc 1974
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2055 gttatttaacttaaccatgatgttgattgaactggaactgactatgagaggtacaacaccaa 2114
OY 1975 gtccctcagcacaatcggggaactcaacgagttctctggtgagctcaagggccggcggtg 2034
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2115 ctgcttttcatTTGTGAGGGATCTGAACCAATACAGACAGACATAAAGAAATGGGCTG 2174
OY 2035 accctgcagtggtgtgactcgagcgaggggaactataccggcgcgcgaaagcgcg 2094
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2175 acttttcaatggcgtgatttctgctgctgagacatttttccgtgcttccacacttaaa 2234
OY 2095 cagagagatctacagctcgagaggaagagacgagcgactgacacgcatgtacaagtgcg 2154
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2235 acagatttgcgaatctgagaaacagacagatttgcattgatagaagaaactcaatgatct 2294
OY 2155 ataagcggtgtgaggttcaacttcttccagtaagtggtgcgaacggaactcccgcttc 2214
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2295 gtcatgagagtgagatgattcttcttcccttaccgtatcttccaaamaagcttcttcc 2354
OY 2215 ccccatctcatgagcggtgagagacacagctgagcggtcggtctgtgacacacggcg 2274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2355 ccacatgcttcttggggctccggctctcacaacgctgacagctttagtgagaaacttgaaa 2414
OY 2275 ctgc 2278
    |||||
DB 2415 ctgc 2418
    |||||

```

```

RESULT 6
HUMTERR HUMTERR 2826 bp mRNA PRI 14-JAN-1995
LOCUS

```

```

DEFINITION Human transferrin receptor mRNA, complete cds.
ACCESSION M11507
VERSION 9339515
GI:339515
KEYWORDS transferrin.
SOURCE Human, cdna to mRNA, clone pCDTR-1.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2826)
AUTHORS McClelland,A., Kuhn,L.C. and Ruddle,F.H.
TITLE The human transferrin receptor gene: genomic organization, and the
complete primary structure of the receptor deduced from a cDNA
sequence
JOURNAL Cell 39 (2 Pt 1), 267-274 (1984)
MEDLINE 85048936
COMMENT Draft entry and clean copy sequence for [1] kindly provided by
A. McClelland, 17-Apr-1985. The area coding for the transmembrane
domain of the transferrin receptor protein is found at positions
279-362.
FEATURES
     source
         1..2826
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="3q26.2-qter"
            <1..>2826
            /note="tftr mRNA"
            96..2378
            /gene="TFRC"
            /note="TFRC"
            /note="transferrin receptor"
            /codon_start=1
            /db_xref="Gene:600-120-433"
            /protein_id="AA61153.1"
            /db_xref="PDB:3J9516"
            /db_xref="GI:339516"
            /translation="MMDQARSFSLFGEPLSYTPRESIANOVQDMSHYEMKLAIVE
            EENDNNKTAANTVTPKRCSSGTCGTAIVYFLIGMIGTLGCKVEPTEBRILA
            GTESPVDEPEDEPFAARLYWDLKKRLSKLSDSTPTSTIKLNNESVYPRRSG
            KDELALAYENQREFKLSKWRDQHEVKIQVDSQNSVLIIVDKNGRLVYLVENPG
            YVAASKATVYGVKLVANFGKRDQEDLYTPVNGSIVIVRAGITFAEKVNAESLNA
            IGLVINDQTEPLVNAELSPFGAHLGTPYPPGPFSEHNGQPPSRSGLEINPY
            OTISRAAEKFLGMEGDQPSMDKTDSTGCRVPSSESNKLYTVNVLKEIKLITGCV
            IKGFVEDHYVYVGAQDPDAMPGAAGVGTALLLKLAKQMSDVLKLDGFPSSITF
            ASWSAGFSGVAGTEWLEGLYSLSHLAFYIINDKAVLGTSNKSVAASPLLYLIEK
            TMQNVKHPVYGOFLYQDSNMAKVEKLTLDNAAPFLAYSGIIPAVSFCFBDYPLY
            GTTQDYKELIERLPELNVKARAAYVAGQFVIRLTHDEVLEINDYERYNSQLSFVD
            LNOYRADIKEMGLSLOWLYSARGDFEATSRITLDFGNAEKTRDFYMKKLMDRYMVE
            YHFLSPVSPRESPEFRHVPFGSSHTLPALLENLKLNKNGAFNETLFRNQALATIV
            TIGQAAALSDVDIDNEF"
BASE COUNT 793 a 568 c 676 g 789 t
ORIGIN 449 bp upstream of PstI site.

```

```

Query Match 9.7% Score 280.2; DB 10; Length 2826;
Best Local Similarity 55.0%; Pred. NO. 1.3e-36;
Matches 717; Conservative 0; Mismatches 348; Indels 38; Gaps 7;

```

```

OY 980 cagtgtatggaatctgacactggaactggaacccctacacacactggtccctctc 1039
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 982 catttcttggacatctctatcgtggggacagggacaccttaccacacttgattcccttct 1041
OY 1040 tcaatcaaacacagcttccctcagttgatacatcagagcttccacgacatccacccagc 1099
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1042 tcaatcacacactcagtttcaccatctcggtgacatcgagattgcctaatattacctgacaga 1101
OY 1100 ccatcagtcagacattgctcccgccgctcgagagaaactcaaggg-ccctgtggtcccc 1158
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1102 caatctccagagctgctcagaaaaagctgttgggaaatggaagagacgtgtccctctg 1161
OY 1159 caagaatgacagggagagcctccttaagctcccttaaccctgggcccggccacagactg 1218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


Db 1162 ACTGAAAAACAGACTCTACATGTAGATGTAACCTCAGAAAACAAGATG-----TG 1214
Oy 1219 cggcagtagtgaacaacacacagaccctccaccctcaacaacatcttgccgctgcatc 1278
Db 1215 AAGCTCACTGTGAGCAATGTGCTGAAAGAGATAAAAATTTCTTAACATCTTTGGACTTAT 1274
Oy 1279 gaagccgcctcagacagatcaactacgtgtgcatcgggcccaagagatcatcggggc 1338
Db 1275 AAGGCTTTGTAGAACACAGATCATTATGTGTAGTTGGGCCACAGAGATCATGGGGC 1334
Oy 1339 ccaggaagcagctaaatccgcgtgtggggaacgactactctcgtgagctgtgctgagac 1398
Db 1335 CCGTAGGCTGCAAAATTCGGGTGTAGGACAGCTCTCTATTGAAACTTGCCAGATGTG 1394
Oy 1399 tccctcagcagc---tgagcaacggcttcggcccgccagagatcctcctcctcagctgg 1455
Db 1395 TCAGATATGCTTAAAGATGGTTTCAAGCCACAGAGATTAATCTTTGCCAGTTGG 1454
Oy 1456 gacggtggtgacttggaagcgtggggtccacagagtgctgaagagctactcagcgtg 1515
Db 1455 AGTGTGAGAGACTTGTGATCGGTGTGTCGCACTGAATGGCTAGAGGATACCTTTCGTC 1514
Oy 1516 ctgcaccccaaaagccgtagtgcagcgtggaacgagcgtgctggggatgacaag 1575
Db 1515 CTGCAATTAAAGGCTTTCACCTATATTAATCTGATTAAGCGGCTCTTGATCCAGCAAC 1574
Oy 1576 ttctctgcagaacacagcccccctctcgaagctcattgaagatgctcctgaagcagtg 1635
Db 1575 TTCAGGCTTTCGCGACGCCACTGTTGATACGCTTATTGAAAAACAATGCAAAATGTG 1634
Oy 1636 gattctcccaacacaggtgggagagactctctatgacaggtgtgtctcaccatccagc 1695
Db 1635 AAGCATCCGGT---TACGTGGCAATTCTATATCAGA-----CAAC 1673
Oy 1686 tggagatgctgagtgatcccgccctcaccatgagcagcagtgctatcctcctcagcc 1755
Db 1674 AACTGGCGCAGCAAGTTGAGAACTCACTTACAAATGCTTCCCTTCCCTTGCA 1733
Oy 1756 ttctggggaagtcctccgctgagtgctcctcttctgaaggaagcagcagcctacatc 1815
Db 1734 TATTTGGAATCCACGAGTTTCTTCTGTTTTCGGA--GACACAGATTATCTTAT 1790
Oy 1816 ctgcacaaaagaggaactatgagaacctgacaaagtgtcgaagccgctgccc 1875
Db 1791 TTGGGTACCACTATGACACCTATATAGGAACTGATGAGAGATTCTG--AGTTGAAC 1847
Oy 1876 gccgtggccagcagccgtgcccagctgcagggcagcagctcctcctcagcagcagat 1935
Db 1848 AAAGTGGCAGCAGAGCTGACAGAGTGGTGTGATGTTGATTAACCTAACCATGAT 1907
Oy 1936 cgcgtcgtccctcgaactcggcgtcaggggaagctcgtcccaagcagcagcagc 1995
Db 1908 GTTGAATTGAACCTGACTATAGAGGTACAACAGCAACTCTTTCATTTGTGAGGAT 1967
Oy 1996 ctcaacagagttctctgggaacctcaagggccgagcgtgacacctcaggggtgactgc 2055
Db 1968 CTGAACCAATACAGACACATTAAGGAATGGGCTGAGTTTACAGTGGCTGATTT 2027
Oy 2056 gcgcgggggagactacatccggcggcgggaaagcgtgcgagcaggaatcctcagccgag 2115
Db 2028 GCTGTGAGACTTCTTCCGCTCTACTTCCAGACTAACACAGATTTGCGGAATGCTGAG 2087
Oy 2116 gagagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2175
Db 2088 AAAAGACAGATTTGTATGAAGAAACCAATGATCGTGTATGATGAGATGAGATAC 2147
Oy 2176 ttctctccagtagtgcgcagcagcagcagcagcagcagcagcagcagcagcagcagc 2235
Db 2148 TTCTCTCTCTCTACGTATCTCCAAAAGAGTCTCTTCCGACATGCTTCTGGGGCTCC 2207
Oy 2236 ggaagcacaagcgtggcgcctctgtgagccacgtgcgctgc 2278
Db 2208 GGCTCTCACAGCTGCGACCTTACTGAGAGAACTTGAACCTGC 2250

RESULT 7
CRUTRANREC CRUTRANREC 2460 bp mRNA ROD 09-DEC-1993
LOCUS Cricetulus griseus transferin receptor mRNA, complete cds.
DEFINITION L19142
ACCESSION L19142
VERSION 9304528
KEYWORDS L19142.1 GI:304528
SOURCE transferin: transferin receptor.
ORGANISM Cricetulus griseus cDNA to mRNA.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
REFERENCE 1 (bases 1 to 2460)
AUTHORS Collawn,D.F., Lai,A., Domingo,D., Fitch,M., Hutton,S. and
Trowbridge,I.S.
YRFR is the conserved internalization signal of the transferin
receptor, and a second YRFR signal at position 31-34 enhances
endocytosis
J. Biol. Chem. 268 (29), 21686-21692 (1993)
FEATURES
MEDLINE
SOURCE Location/Qualifiers
1..2460
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/cell_line="CHO-K1"
/tissue_type="Ovary"
/tissue_lib="Stratagene catalog #937910"
1..2274
/codon_start=1
/function="endocytosis"
/product="transferin receptor"
/protein_id="AA03576.1"
/db_xref="PID:g304529"
/db_xref="GI:304529"
/translation="MMDQARSAISNLFGEPLSYTRPSLIAROVNDGSHVEMKLAVDE
EENTDNMKASVKKRRLNGRLCGTIAVYIFLIGMIGLIGCKRTBOCDYRLAE
TETGENSELIDENIPQSSRLVWADLKLLEKDLAIEFTDIKOLDSOTSRAGSKDE
NLAYIENQFDEFLISVWREDEHYVKIOVGSAAONAVTIINVGDSDLVNPQGYVA
YSKATVSGKLIIHNFGLKDFEDLKPVNSLIVRAGKITPEKVAANQSFVAIG
LIVMDOTKPPVVEALSLFGHAHGLGTDPTPGPSPFNHQPSPSSGGLPSIVORT
SRKAELKPNMFCNCPNMTDCTLESQSGGINVLSVNNVKRPIILNFEVIG
FEEDRITVGAQDANGPGNAKSVCTGLLKLQAFSDVNSGGRPSISIFASV
SAGPGVAGATEMEGLSLSLHLKAFYINLKVAGLDRNKVSAASPLLYLIEKTMQ
DVRHPIDGKPLVRDSNMIISKVEDLSLNAAFPLFVAGIPVASFPCENEDYPLDTN
LDYREKLQKVPOLNKRMAAEVAGQFIKFLHDIELNDYDMYNNKILISVELNO
FRADIKAMGSLQWLSYARGDFPATSRLITDFPNARKTNFVYRATINRMKVEYHF
LSRIVSPRESRFRIIFPGSGSHLITALEVNLKLNKNSAFNETLFRQALATWTTIO
GVANALSGDILWDIDNEF"
BASE COUNT 744 a 487 c 539 g 690 t
ORIGIN
Query Match 8.9%; Score 256; DB 12; Length 2460;
Best Local Similarity 54.2%; Pred. No. 1,16-32;
Matches 713; Conservative 0; Mismatches 565; Indels 38; Gaps 8;
Oy 980 cagtgatggaatctggaactggagagcctcaacacactgcttcctcct 1039
Db 878 CACTCTTTGGACATCTCATCTAGGACCTGGGATGCCATACACTGGTTCCATCTT 937
Oy 1040 tcaatcaaacacagttcctcagctgcatcagcagcctccagcagcagc 1099
Db 938 TCAATCATACACAGTTCTCCATCTCAGTCAAGATTAACCTGTTGACGA 997
Oy 1100 ccattatgacagacattgctcctccgcctgcgaggaagct- caaagccctgtgcccc 1158
Db 998 CGATCTCAAGAAAGGCTCAGAAAAGACTATTTCAAAACATGGAACAACTGCTCCTTA 1057
Oy 1159 caagaatgscagggagagcctcctagcctccctatcagcccgggcccgccagcagatg 1218

```

Db 1058 GTTGACACAGATTTCTTGTAGCTAGTAATCTCA-----CAGGATATAATGTG 1110
QY 1219 cggtagtgcgaacaatacaagaactcccccatacaacatcttcgctgcatc 1278
Db 1111 AACCTCTGTGAAACAGTGTCTGAAAGAAACAAGAAATTAACATCTTTGGAGTTAT 1170
QY 1279 gaagcgctcagagcagatcactacgtgtgtcagggcgccagagagatgcattgggac 1338
Db 1171 AAAGATTGTGAGAACAGACGAGCGGTACATTAGTAGTGGCCAGAGAGATGCTGGGGT 1230
QY 1339 ccagagcagactaaatccgctgtgtggagcggctactcctgtgagctgtgcgacctt 1398
Db 1231 CTTGCTCTGTGAAATCAGTGTGGAAACAGGCTTCTGTGTAACCTTGCCCAAGCATTC 1290
QY 1399 tcccacatgfr---gagcaacggtcccgcccgaggaagtcctcccttataagcttg 1455
Db 1291 TCAGATATGTTTCAAGAGGTGGGTTTAAACCCAGAGAACATATCTTTGGCAGCTGG 1350
QY 1456 gacggtgtgcttcttggaagcgtgtgcctcagagagtgagagagctaccagcgtg 1515
Db 1351 AGTCAGAGAGACTTCGAGAGCTGTGTGTCACAGAGTGCTAGAGGATTAACCTTCACT 1410
QY 1516 ctgacactcaaaagcagtagtgcgtgagccttgagcaacgagctgtggtggatgacaag 1575
Db 1411 CTGCATTGGAAGCTTCACTTATATTAATCGATTAAGTTTCCTGTGTAAGAAC 1470
QY 1576 ttcatgccaaagacacgccccctctctgacaagcttcattgagagtgctccgtgaagcagtg 1635
Db 1471 TTCAGGTTTCTGACCCCTCTATATATACACTTATGGAAGAACATCCAGAGCGTA 1530
QY 1636 gatctcccaacacagtgagcagactcctatgaacagsgtgcgttcaaccaatccagc 1695
Db 1531 AGAATTCGAATGTA---TGGAATACCTCTCTATGAGAC-----CAGAAATGGATC 1578
QY 1696 tggagtcgtgaagtgtacccggccctacccaatgacagcagtgctcttccctcaagcgc 1755
Db 1579 AGCAAACTGTGAGG-----ATCTTTCCCTTGACAAATGCTGCTTCCCTTTTGCA 1629
QY 1756 ttgtggagtgccctgcgtgagtttcccttattgagagcagcagcgtctaccatctc 1815
Db 1630 TATCTGGAATCCAGAGAGTTCTTCTGTGTTGTAGAGATG---AGGATATACCTTAT 1666
QY 1815 ctgcacacaaagagagacactatgagaacctgcataaagtgctgcacaagcgccctccc 1875
Db 1687 TTGAGACACTATTTGATTAATCTAGAGAACTAATTCAGAAAGTTC---CTCAGCTGAAC 1743
QY 1876 gcgctggccaagcgctgtgcccagctgcagggcagctcctcatcccgctcagcagcat 1935
Db 1744 AAAATGTTGTGTCAGCAGCAGAGAGTAGTGTGCTGATTAATAAATTAACTTACCAGAC 1803
QY 1936 cgctgtgtccctgcagcttcgcgctcagggagagtgctccctcagggcaatcgggagc 1995
Db 1804 ATTGAATCTGAACCTGTGACTATGATATATMAACAAGATATGTCATTTGTGAAGGAG 1863
QY 1996 ctcaacagatctctctgagagcctcaagggcccgagctgcagcagctgagtggtactgc 2055
Db 1864 CTGAGCAATTCACAGAGAGATATCAAGCAATGGGCTGTGAGTCTACAGTGTCTATTTCT 1923
QY 2056 gcgcgggggagactacatccgagcgggcgaaagactgcgagcagagatctaaagctggag 2115
Db 1924 GCTGTGTGAGACTTTTCCGCTACATCTAGACTGACGACTGATTCATTAATAGCTGAG 1983
QY 2116 gagggagagagagcagcagcagcatgtlatacagtgcgataatgcggggtggagttctac 2175
Db 1984 AAAACAACAGATTTGCTGTAGGGAATCAATAACCGTATTAAGAAAGTGAATATAC 2043
QY 2176 ttcccttcacagtagtgcgcagcagcagctcccggttcgcgacatctcatgagcgt 2235
Db 2044 TTCTGTGACCCCTACGATATCTCCAGAGAGTCTCTTTCCAGACATCTTCTGGGGTTC 2103
QY 2236 ggaagacacagcgtggcgccctgtgtgacacactgcgagctgtgcgtcacaagc 2291
Db 2104 GGCTCTACACTCTACGGCTTTAGTGAGAACTTGAAGCTTCTGTCAAGAAACAG 2159

```

```

RESULT 8
RATFR 3413 bp mRNA ROD 27-APR-1993
LOCUS Rat transferrin receptor mRNA, 3' end.
DEFINITION M58040
ACCESSION M58040
NID 9207463
VERSION M58040.1 GI:207463
KEYWORDS transferrin receptor.
SOURCE Rat adult 120 day old testis, cDNA to mRNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 3413)
AUTHORS Roberts,K.P. and Griswold,M.D.
TITLE Characterization of rat transferrin receptor cDNA: The regulation
of transferrin receptor mRNA in testes in Sertoli cells in culture
Mol. Cell. Endocrinol. 14, 531-542 (1990)
FEATURES
Source
1..3413
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/def_stage="120 day old adult"
/rissue_type="testis"
1..1870
/gene="transferrin receptor"
<1..1870
/product="transferrin receptor"
/gene="transferrin receptor"
/codon_start=2
/protein_id="AAA42273.1"
/db_xref="PIR:G207464"
/db_xref="GI:207464"
/translation="TEFTDILKQLSQNTYTPREAGSOKDENLAVYIENLHFDKFSKY
WRDEHYVKIOVNSVSONLVTINSNSIDPVAPRGVAFVSAGETGKLVANSGTK
KPEELNLYVNSGLIVVRAGKITPAEKVANAQSPNAGIVLYMDRNTPEVADLOF
GHAHGTGDPYTPGPFSPFHPNPFPSOSGLSIYVORTSRAPEKLFNMGNCPSP
WNIDSSCKLELSQNONVKLTNNVNLKETIILIFVYIKTEEPDRIVVGAORDMAGP
GVAKSSVGTGLLKLAAOVFSDMSIDGEPSPSIIIFASWTADYGAVGETEVLDEGLS
SLHLAFYIINDKLVLTGINSFVNSPILYLMGKIMQDVHPIDGKLYLNSNWS
KIEELSDNAAPFLAVSGIPAVSFCEDEDPYLGKLDYELIOLKPOLNOMVR
TAAEVAOGFIITLTDIELTLEINFWKYNLSKLSFMDLNOFRADIDMGISLOMYSAR
GVFPRATSLTIDFHNAEKTRFVAREIRDKYDHLNLSVYSPRESPFHITMS
GSHTSALVENLRLRKNITTAFTNELFRNQLALATWITGIVANALSGDIWNIDNEP"
misc_feature
2846..3007
/note="putative iron regulatory region"
misc_feature
3184..3301
/note="putative iron regulatory region"
BASE COUNT 971 a 668 c 741 g 1033 t
ORIGIN
Query Match 8.6%; Score 247; DB 12; Length 3413;
Best local Similarity 52.6%; Pred. No. 2.9e-31;
Matches 767; Conservative 0; Mismatches 625; Indels 65; Gaps 8;
QY 981 agtgcagcagctgtgcaactggaactggaacccctacacacctggtccctcctt 1040
Db 475 ATTCTTGGACATGCTATCTAGAGACTGGGATCATATACACTGGCTTCTCTT 534
QY 1041 caatcaaccacattccctcagctgcatcatcaggcctccagacatccagccagcc 1100
Db 535 CAACCAATACATCTTCCGCATCTCAGTCACTTGATTCCTTCTATATGTCGAC 594
QY 1101 catcagtgacagattgcccctccgctgtcgtgaggaagctaaagc---ccgtgtgcccc 1159
Db 595 GATCTCAAGAGCTCTCTGCAAAAAGCTATTCAAAACATGGAAGGAAGAACTGCTCTAG 654
QY 1160 aagaatggcagggagcctctcagctcccttatacactgagcggcgccagcagctgc 1219
Db 655 TTGGAATATATGATTCCTCATGTAAGCTGGAACCTTCA-----CAGAAATCAAAATGTGA 707

```


| | | | |
|----|------|------------------------------------------------------------------|------|
| QY | 1220 | gacagtagtgcacaaactcaaggactccacccccaataaacaactcttgctgcacacg | 1279 |
| Db | 708 | AGCTCACTGTGAACCAATGTAAGTAAGAAACAAGAACTTAACTTCTTGTGGCTTATTA | 767 |
| QY | 1280 | aagagccctcaaggagccagatcatcactagcttgctatccgggcccagaagatbcaggggccc | 1339 |
| Db | 768 | AAGGCTATGAGGAACCCAGACGCGCTACATTTGTAGTAGAGCCACAGAGAGCGCTTGGGGCC | 827 |
| QY | 1340 | caggagcagctaaatccgcctgttgggagcggtatactccctgagcttggttggagcccttt | 1399 |
| Db | 828 | CTGGTGTGGGAAGTCCAGTGTGGGAACAGGCTCTGTGTGAACCTTCCCAATATTCT | 887 |
| QY | 1400 | ccctccatgct---gagcaaggtctccggcccccgagaagctcctcttcaacgctggg | 1456 |
| Db | 888 | CAGATATGATTTCAAAAGATGATTTTAGCCACGAGAGGATTAATCTTTGCCAGCTGGA | 947 |
| QY | 1457 | acgagtgatgactcttggaagcgtgtggctccagcaggtgtgctaaagagctaccacgctgc | 1516 |
| Db | 948 | CTGACGAGAGACTATGAGCGTGTGGCTCCACACTGATGGCTGAGGGGTAACCTTTCACTT | 1007 |
| QY | 1517 | tgcacctcaagccgtagtgtacgttgagccttggacaacgacgtgtctggggatgtacaagt | 1576 |
| Db | 1008 | TGCATCTAAAGCGTTTCACCTTACATTAATCTGGATAAAGTCCTCTGGGTAAGCACT | 1067 |
| QY | 1577 | tctatgcagaagcagcgcctccctcttgcacaagctcatatggagagtgctcttgaagaagtg | 1636 |
| Db | 1068 | TCAAGTGTTCGCCAGCGCCCTCTATTAATACATTATGGGGAAGATAATGAGAGAGCTAA | 1127 |
| QY | 1637 | attctcccaacacaaagtgggcagagcctctatgaaacagtggtgttaaccaatccacgct | 1696 |
| Db | 1128 | AGCATCCGATTGA---TGGAAATATCTATATTCGAAA-----CAGTA | 1166 |
| QY | 1697 | ggagatgctgagtgatccgcgcctccacatggacaagcagtgctatctcttccacggcct | 1756 |
| Db | 1167 | ATTGATATTACCAAAATATGAGGAACCTTCTGTGGACAATGCTCATTCCTTTCTTGAT | 1226 |
| QY | 1757 | tctgtggagatccctctgcctgcagatcttcccttatggagagcaccaggtcctaccatcc | 1816 |
| Db | 1227 | ATTTCAGGAATCCAGCAGTTCCTTCTTCTGTTTGTGAGAGT---AGGACATATCTTAT | 1283 |
| QY | 1817 | tgcacacaagggaggaacactatgaaacactgataagtggtgtgtaagcgcgctctgcgc | 1876 |
| Db | 1284 | TGGGCACTTAACATGATACCTATGAGATATTTAATTCAGAAAGTT--CTAGCTCAAC | 1340 |
| QY | 1877 | ccggtggccacggcggttggcccagctctgcagggcagctcttccatccgctccagcagatc | 1936 |
| Db | 1341 | AAATGTTTCATGACGACGACGAGAGGTGGCGGTCACTTATTTAAATTACCATGACA | 1400 |
| QY | 1937 | gctctgctccctccactgcgcgcgtacaggggagcgtctcctcagcacatcgggaacc | 1996 |
| Db | 1401 | TTTGAGTTGACCTCGAGACTATGATGATTAACAACGAAACTACTGTATTTAGGAAGATC | 1460 |
| QY | 1997 | tcaacaggtctctcgggagacttaagcccgggccttgaccccttgtaagtggtgtatccgg | 2056 |
| Db | 1461 | TGAACCAAGTTCAANGCAGATATATAAGATATGGGTCTAAGCTTAACAAGGCTGATTTCTG | 1520 |
| QY | 2057 | cgcgggggagactacatcccggtcgcgcggaagaagctgcggaagagatctacagcctcgagg | 2116 |
| Db | 1521 | CTCTGGAGACTACTTCCGTGCTACTTCTTAAGACTAACACTATTTTCATATAGCTGAGA | 1580 |
| QY | 2117 | agagagacgagcgacttgcacacgcatgtacaacgtgcgcataatgcgggttgagttact | 2176 |
| Db | 1581 | AAACAAACAGATTTCGTATGATGAGGAATCAATGATGTAATTTGAAGATGGAGATACAT | 1640 |
| QY | 2177 | tcccttccagtagtgcgtgcacgcccagactccccgtctccgcaactcttcatgggcgtg | 2236 |
| Db | 1641 | TCTCTGTAACCTATGATATCTCCAAAGAGATCTCTTTCGGAACATCTTCTGGGCGCTCG | 1700 |
| QY | 2237 | gagacacacgcttggcgccctctgtgacacaccttggcgtgtgtctgcacaaacgctcgc | 2296 |
| Db | 1701 | GCTCTCACACCTCTCAGCTTTATGTGAGAGAACTGAGACTTGTCTCAGAAAAATAT---- | 1756 |
| QY | 2297 | ggaccccccgggacactcctcctacatggtcttcagagagccggttcttcggcgtcagctag | 2356 |

[illegible]

```

Db 947 CTTCTTCAATCACTACTAGTTTCCGCCAATCTCAGTATCAGGCTGCTTAATATACCTG 1006
Qy 1094 cccagcccaatcagtcagacacatgctcccgctgctgaggaagcctaaag-ccctgtg 1152
Db 1007 TGCACAAACATCTCACAAGCTGTGCGAGAAAGAGCTATTTGGAAAAATGAGAGAGCTGTG 1066
Qy 1153 gcccccacaagaatgagcaggcgagccctcctaggtcccttatccatcctgagccccggcca 1212
Db 1067 CTGCTATGATGAGACATAGATTTCTTCTATGTAAGCTGGAACCTTCA-----CAGAAATCA 1119
Qy 1213 cgaactgcygctagtgtgtcaacaatcagagacctccaccaccatcaacaacatcttcggc 1272
Db 1120 AATGTGAACCTCATGTGTGAAAAACGTAAGAAAGAAAGAAATACTTAAATCATCTTTGGA 1179
Qy 1273 tgcatacgaagccgctcagaagcagatacactacgtgtcatcggggccagagagatgca 1332
Db 1180 GTTATTAAAGGTTATGAGAACCCAGCCGTATGTTGTATGATGAGACCCAGAGAGAGCT 1239
Qy 1333 tggggcccgag--agcagataatccgctgtgaggagagcctactccttgagcgtgtg 1389
Db 1240 TTGGGTCGTGCTGTGCTGCGCGAAGTCCAGTGTGGAAAGGTCCTTTGTTGAACCTTGC 1299
Qy 1390 cggagaccttctcctcctcctggt--gagcaagcgttccggcccccgagaagctcctctc 1446
Db 1300 CAAGTATCTCAGATATGATTTCAAAAGATGATTTAGACCCAGAGATATATCTTT 1359
Qy 1447 atcagctggaagcgtgtgacttggaaagcgtgtgctcagagagtgctagaagctac 1506
Db 1360 GCCAGCTGAGCTGACAGCCACTTGGAGCTGTGGTGTGCTGAGCTGAGTGTGGAGGATAC 1419
Qy 1507 ctcaagcgtcagcctcagcctcaagcgtgagtgagcgtgagcctgagcaagcagcgtcg 1566
Db 1420 CTTTCAATCTTTGCTTTAAAGCTTTCACTTATATTAATTTGATTAAGTTGTCCTTGGT 1479
Qy 1567 gatlgaagaatlcaatgccaagaccacccctctgacaagctcattgagagtgctcg 1626
Db 1480 ACTAGTAACCTTCAAAAGTTCTGCGACCCCTTATATATACATTAATGGAAGAAATAG 1539
Qy 1627 aagcaggtgagatctcccaaccaagctgagcagcctctatgaaagagtggtgtcacc 1686
Db 1540 CAGAGATGTAAA-----GCATCCAGTGTGATGAAAAATCTATATATAG-----ACAGC 1587
Qy 1687 aatcccaagctgagatgctagtgatgcgccccctaccatgagcagcagtgctcctacc 1746
Db 1588 AATGTGATTAGCAAAATGTAGAAA-----CTTCTTTGACAAATGCTGATATCTCT 1638
Qy 1747 ttcacggccttgtgagagtcctcgcgtcgagttctccttatgagagcagcagcc 1806
Db 1639 TTCTTGTGATTTGGAATCCGACAGTTCTTTTGTGTTGTGAGGATG---CAGAC 1695
Qy 1807 taccatctcgcacacaagaagagacactatgaaagcctgacataagtgctgcagagc 1866
Db 1696 TATCTTATTTGGGCACTGATGTGATGATCTTATGAGCATTTGCTAGAAAGTTC---CT 1752
Qy 1867 cgcctgcgcgcgcgtgcccagcagcgtgcccagcgtcagagggagctctcctccggtc 1926
Db 1753 CAGCTCAACCAATGTTGTGTACAGCAGGAGAGTGGCTGTGACGTCATTATTAACCTT 1812
Qy 1927 agcagatcgcctgctccctcgcactcgcgtcgagcgtcagggagcgtcgtcctcagcac 1986
Db 1813 ACCCATGAGCTGATTTGAACCTGAGCTATGAGATGTATTAACAGCAAACTACTGATATT 1872
Qy 1987 atcgggaacctcaagagttctctgaggacctcaagcccgcggtcgtgacctgcaagtg 2046
Db 1873 ATGAAGGATCTGACACAGTTTCAAAACAGATATCAGGAGATATGAGTATAGTCTACAG 1932
Qy 2047 ggtgactcgcgcgggggagactacatcagcggcgaggaagcgtgagagagatctac 2106
Db 1933 CTGTATTTCCGCTCGTGGAAGACTACTTCGCTGCTACTTTAGACATAACAAGTATTTTCA 1992
Qy 2107 agctcggagagagagcagcagcgtgacacgcattgaaagcgtgagcctaaatgcyggtg 2166

```

```

Db 1993 AATGCTGAGAAAACAACAGATTTGTCATGAGGGAATCAATGATGATTAATGAAAGTG 2052
Qy 2167 gagttctactctcttcccaagtagtgcgcagcagcactcccgltccgcacatctc 2226
Db 2053 GAGTATCACTTCTCTCCGCCCTATGATATCTCCAAAGAGTGTCTCTTCGATATCTTC 2112
Qy 2227 atgggcgtgagacacagcgtgagcccgctcgtgagcagcctcgcgtgtggtcc 2286
Db 2113 TGGGCTCTGGCTCTCACACTCTCTACACTTTAGTGGGAACCTTGAACCTTGTCAAAA 2172
Qy 2287 aacagctccggagcccccggggccactctccactgagcttccagagagccgltccgg 2346
Db 2173 AATATT-----ACTGCTTTAATGAACCCCTTCAG 2205
Qy 2347 cgtcagcagccctgctcactcgtgagcgtgcagagggcagccaatgcttaggggat 2406
Db 2206 AACCAATTGGCCCTGGCTACTGTGACTATTCAGGAGATGCAAAAGCCCTCTCTGTGAC 2265
Qy 2407 gctggaacattgataacaactctga 2433
Db 2266 ATTGGAATATTGACATGATGTTTAA 2292

```

```

RESULT 10
LOCUS 123809 780 bp DNA PAT 07-OCT-1996
DEFINITION Sequence 27 from patent US 5538866.
ACCESSION 123809
VERSION 91603679
KEYWORDS NID
SOURCE 123809.1 GI:1603679
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 780)
AUTHORS Israeli,I.R.S, Heston,W.D.W. and Fair,W.R.
TITLE Prostate-specific membrane antigen
JOURNAL Patent: US 5538866-A 27 23-JUL-1996;
FEATURES
Source Location/Qualifiers
BASE COUNT 197 a 182 c 201 g 176 t 24 others
ORIGIN

```

```

Query Match 5.6%; Score 160.2; DB 5; Length 780;
Best Local Similarity 57.1%; Pred. No. 3.9e-17;
Matches 391; Conservative 0; Mismatches 265; Indels 29; Gaps 5;

Qy 988 ggaatgtagcactgaggagactggaagccctacacacagcctgctccctcctcaatcaa 1047
Db 17 GGACATGGCCACCTTGAGACTGGAGACCCCTTACACCCAGAGCTTCCCTTGTCACACAC 76
Qy 1048 accca-----gtccctccagttgacatcacagcgtctccacagcatccagccagccatc 1104
Db 77 ACCCANNNGTTTCCACAGTTGATCTTCAAGACTACACCCACATGCTGTTCAGAGCATC 136
Qy 1105 agtgcagacattgctcccgctcgtgaggaagctcaag--gccctgtggcccccaag 1162
Db 137 TCTACAGTGCAGCAGCCAGGCTGTTCAGCAAAATGATGAGACATGCTGTGANAGN 196
Qy 1163 aatgcaaggagagcctctaggtcctccttatcaactg-----ggc 1203
Db 197 NGTTGAAAGGTGCATCCANNNTTCTGTAGGTGNNACNNNAACAAGAGAGANNNN 256
Qy 1204 cccgggcaagcagctggcgtagtggtcaacaatcagagacctccaccatcaaacac 1263
Db 257 GCCAANTAAATGTTAAACTAGATGTGAACAATTCATGAAAGACAGAAATTTGTGAAC 316
Qy 1264 atctcgtcgtcatcgaagccgcgtcagagccagatcactacgtgtgcatcggggccag 1323
Db 317 ATCTCGTGTCATCAGAGGATTTGAAGAACCTGATCGGTATGTTGATTTGAGACCCAG 376

```

| | | | |
|----|------|---------------------------------------------------------------|------|
| Oy | 1324 | agggatgcatggggcccaaggaagcaaaatccgcctgctggagcgtatactactctgag | 1383 |
| Db | 377 | AGAACTCTCTGGGGCCCGAGAGTGGCTAAAGCTGGGACATGGAACTGCATATTTGTTGG | 436 |
| Oy | 1384 | ctgctggcgacaccttcctccatcgtgtgaagcaac---gacctccggcccgcaagatctc | 1440 |
| Db | 437 | CTTCCCGGTGTGATCTCAGCATATGTAATAAAGAGAGGCTACAAACCGAGCGAAGCATC | 496 |
| Oy | 1441 | ctctctacatcgctggagcagctggctgtgacttggaaacgctggctccacgagtgctaa | 1500 |
| Db | 497 | ATCTTCTGCTAGCTGAGACTGCAGAGACTACGGAGCTGTGGTGTACTAATAGCGCTGAG | 556 |
| Oy | 1501 | ggctacctcagcgtctgctgacccctcaagaccgtatgctgctga--gctcggacaacgac | 1558 |
| Db | 557 | GGGATCTCTGCATCTGCTGATGCCAAAGCTTCACTTACATCANNCGTTGGATGCTCAG | 616 |
| Oy | 1559 | tgcctggggatgtgaaagtcttcatgccaagaaccccccctcttgacaagtctcatatga | 1618 |
| Db | 617 | TCCGCGGAGGAGGACGATGTAAGATTTCTGCCCCAGCCCTCTGTATATGCTGTGGGGA | 676 |
| Oy | 1619 | gtgcctgaagcaagctgcatctcc | 1643 |
| Db | 677 | GTAATTGAAAGGGGTGAGAAATCC | 701 |

| | | | | | |
|------------|-----------------------|-----------------------|-----|-----|-------------|
| RESULT | 11 | | | | |
| GI4981/c | | | | | 04-JAN-1996 |
| LOCUS | GI4981 | 250 bp | DNA | STS | |
| DEFINITION | human STS SHGC-14854. | | | | |
| ACCESSION | GI4981 | | | | |
| NID | 91131744 | | | | |
| VERSION | GI4981.1 | GI:1131744 | | | |
| KEYWORDS | STS sequence; primer; | sequence tagged site, | | | |
| SOURCE | human. | | | | |

REFERENCE
AUTHORS
JOURNAL
1 (bases 1 to 250)
Myers, R.M.
Unpublished (1995)

Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu

Primer A: CTACACAGCTGGCAGAGTCC
Primer B: AGCTGACCTCATGTGACGG
STS size: 139
PCR profile:

| | |
|---------------------|-----------------------------|
| Initial incubation: | 94 degrees C for 90 seconds |
| Denaturation: | 94 degrees C for 15 seconds |
| Annealing: | 64 degrees C for 23 seconds |
| Polymerization: | 72 degrees C for 30 seconds |
| PCR Cycles: | 30 |
| Thermal Cycler: | Perkin Elmer 9600 |
| Protocol: | |
| Template: | 25 ng |
| Primer: | each 1 uM |
| dNTPs: | each 200 uM |
| Tag Polymerase: | 0.05 units/uL |
| Total Vol: | 10 uL |

| | |
|---------------------|--------|
| Buffer: | |
| MgCl ₂ : | 2.5 mM |
| KCl: | 50 mM |
| Tris-HCl: | 20 mM |

| | PH: | 8.3 |
|-------------|---------------------------------------------------------------|----------|
| FEATURES | Prepared with primer pairs derived from T64983--Merck/UniEST. | |
| source | Location/Qualifiers | |
| | 1. .250 | |
| STS | /organism="Homo sapiens" | |
| primer_bind | 99. .237 | |
| primer_bind | 99. .118 | |
| | complement(218. .237) | |
| BASE COUNT | 60 a 60 c 67 g 59 t | 4 others |
| ORIGIN | | |

Query Match 5.1%; Score 146; DB 13; Length 250;
Best Local Similarity 83.3%; Pred. No. 9.6e-15;
Matches 214; Conservative 0; Mismatches 33; Indels 10; Gaps 4;

| Accession | Sequence | Length |
|-----------|-----------------------------------------------------------------------|--------|
| Oy | ccctgagcgtgtagctaacccctaatgagacggtcacatcgtggttaactcaagagagtagcat | 2673 |
| Oy | 2614 ccctgagcgtgtagctaacccctaatgagacggtcacatcgtggttaactcaagagagtagcat | 2673 |
| Db | 247 CCTGAGGTGAGTACCTATGTGAGG---CATACTGGGTATATAGAGAGTAGCAT | 191 |
| Oy | 2674 cccttcaatcacagcccttcctccttcttggtgttcctccatactagagaccactctggg | 2733 |
| Db | 130 CCGTTCATTCACAGCCCCCTTCCC--TTTCTGAGGGCTTCATACCTAGAGACCACCTCTGGG | 133 |

QY 2734 aggttgctaagccctgtagacctgtagccagctctgttagtggtagagatcgcctggcaccat 27933
||||||| | | ||||||||| ||| ||||||||| |

QY 2794 agccttatggccaacagtggtctgtgtgaaggcggtggagttccaataatcaataa 2853

Db 77 TACCTTATGCCCAACAGGTGGCNTGTGGTGAAGAAGGGCGTGGAGTTTCAATATCATATAA 18

QY 2854 ccacctgatatacaataa 2870

Db 17 CCACCTGATATCTATAA 1

| | | |
|------------|-------------------------------------|-------------|
| RESULT | 12 | |
| 123811 | | |
| LOCUS | 123811 | 540 bp |
| DEFINITION | Sequence 29 from patent US 5538866. | DNA |
| | | PAT |
| | | 07-OCT-1996 |

```
VERSION      I23811.1  GI:1603681
KEYWORDS
SOURCE       Unknown.
```

UNKNOWN;
Unclassified.

| | |
|---------|---------------------------------------------|
| AUTHORS | Israel, R.S., Heston, W.D.W. and Fair, W.R. |
| TITLE | Prostate-specific membrane antigen |
| JOURNAL | Patent: US 5538666-A 29 23-JUL-1996; |

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .540 |

| | | | | | |
|------------|-------|-------|-------|-------|----------|
| BASE COUNT | 152 a | 106 c | 133 g | 147 t | 2 others |
| ORIGIN | | | | | |

| | | | | |
|-----------------------|-------|----------------|----------|------------|
| Query Match | 4.48 | Score 127 | DB 5 | Length 540 |
| Best Local Similarity | 60.48 | Pred. No. 9 | Ce-12 | |
| Matches 261 | 0 | Mismatches 166 | Indels 5 | Gaps 3 |

Qy 1217 tgcgctagtgtgcaacaatcacagacctccacccccatcaacaacatttcggtgca 1276

DB / 5 TGAAGCTCACCTGAGGCAGTGTGCTGGAAGAGATAAATAATCTTACACATCTTGGAGTTA 13

12/1 tcgaagccgcctcagagccagatcactacgttgtcattcggggccagagggatgcatggg 1336

Db 135 TTAAAGGCTTTGTAGAACAGATCCTATGTGTAGTTGGGGCCAGAGAGATGCATGGG 19

QY 1337 gccagagcagctaatac-cgctgtgggacggtatactcctggaagctgtgacgacc 1395

ACCESSION AJ012371
NID 94539526
VERSION AJ012371.1 GI:4539526
KEYWORDS naaladase L gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2320)
REFERENCE Pangalos,M.N., Neefs,J.M., Somers,M., Verhasselt,P., Bekkers,M.,
AUTHORS van der Helm,L., Fraiponts,E., Ashton,D. and Gordon,R.D.
TITLE Isolation and expression of novel human glutamate carboxypeptidases
with N-acetylated alpha-linked acidic dipeptidase and dipeptidyl
peptidase IV activity
JOURNAL J. Biol. Chem. 274 (13), 8470-8483 (1999)
MEDLINE 99185063
REFERENCE 2 (bases 1 to 2320)
AUTHORS Pangalos,M.N.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) Pangalos M.N., Biotechnology, Janssen
Pharmaceutica, 30 Turnhoutseweg, 2340 Beerse, BELGIUM
FEATURES
source
1. .2320
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q12"
/tissue_type="small intestine"
17. .2239
/gene="naaladase L"
17. .2239
/gene="naaladase L"
/function="peptidase"
/evidence="experimental"
/product="NAALADASE L protein"
/protein_id="CA39968.1"
/db_xref="PID:e1420518"
/db_xref="PID:g4539527"
/db_xref="GI:4539527"
/translation="MONTKVLGLGLGMAALLGILGLHFAIPKKNASLAPODLLEI
LETVMGLDAHRIRENLRELSREPLASSPRDDLYOLLRKDESGDSSEATY
EVLVSFPOEPNVVDVINGPTGIIHSCRTREENVGDDGPPVQPYAPASCTPO
GLLYANRBAEDFKELQTOGILECTIALTRIGYGRKAKAVNAKHGAVGLVYTD
PAINDGLSPDETFPNSWYLPSPGVKSYIEFGDPLPYAPVSSFRVLANVS
GPEPIPTQIGQDARDLCLNGLTAPMTWQALCHYRLGFGFDDFPDSCVN
VSYNRLERLNSNVGLIRGAVEPDRVLYGNHRDSWVGAVDPSSGAVLELSRV
LGLTKGTMPRRSIVFASMGAEFGIGSTFEFEPFNKLOERTVAVIYNDISVEA
NATLRVQGTTPVQSVFSAKKEIRSPGPDLSIYDMWIKYFNSSPVYGLVPSLSLG
AGSDAPFYHFLGISMDIATITDKSKIRLPTVHTAFDITFDYDKLDPFSSHO
AVALTAGSVILRLSDSFLPLKVSYSFETLRSLQAQODGLLBOHSLSLQPLVTA
VEKFEAALAGRISTLQKSPDLPLOVRMLNDQMLLETLNPRAFPEERYSHVL
WAFSHLRSHIPELISNACSRARDTASGSEMAVEVQRLSIVTALBGAATLAPVADL
"

BASE COUNT 453 a 725 c 671 g 470 t 1 others
ORIGIN

Query Match 3.2%; Score 91.6; DB 10; Length 2320;
Best Local Similarity 46.6%; Pred. No. 3.7e-06;
Matches 672; Conservative 0; Mismatches 674; Indels 96; Gaps 8;

QY 672 tgaagccggaagagtcgagagcagctgcgctgaggaaccctgaactctactgccccta 731
DB 415 tgaaggaanaagctgacccggagcaagggggccagatggtgacaaacctatgctgctta 474
QY 732 cagcgcatcggcaagtcacgagagagctggtgtacgcccacacgaaggcgcccgaga 791
DB 475 tgcctcttcttggaaccccaacagggcctctctgtatgccaacggggcgaggagaga 534
QY 792 cctgcagagcctgcggcgcaagcgtgatccagctgagccgctgctgctgctgctgct 851
DB 535 ctttaagagactacagactcagggcattcaacttgaaaggcacactttgcccctgactcgata 594

QY 852 gggggtgatcagcttcgcccagaaggtgaccaatgctcagacttcggggctcaagagact 911
DB 595 tgggggtgtagggcgtggggccaaagcctgtgaacgtgccaaacagggggttacctgggg 654
QY 912 gctcatataccagagccaagcagactctccaggagcccaacccaagccaagcctgtccag 971
DB 655 gctggtgtacacagacccttgcgcacatcaacgattgggctgagctcacccagcaaacctt 714
QY 972 ccagcagcagctga-----tgacatgtgaccttgggaacttggagagc 1014
DB 715 tcccaactcctgtactctgccccctcagagagtgagagcgagcgtctactacagagatttt 774
QY 1015 ccccaacacactggtctccctcctcctcaatcaaacccagttcctccatgctatc----- 1071
DB 775 tggggaccctctgacttccctacacttccagccgtccctcttcccttccgctggacccttgc 834
QY 1071 -----atcaggccttcocccagcatcccaagcccaatcagtgcaagacttccctcccg 1124
DB 835 caattgctccggatttccccaattcccttacacagaccattgcttccagatccaaagaga 894
QY 1125 cctgcctgaagaaagcctcaaaagcctgttgcccccaagaattggcagagggagcctcctag 1184
DB 895 cctgctctgtaacactcaaggaactttggcccgacacacttgacag-----gaccactggg 951
QY 1185 ctcccctatcaccttgggcccgggccaagac----- 1217
DB 952 ctccactacagagctgtgcttccgcttccgcttccgcttccgcttccgcttccgcttccg 1011
QY 1217 --tgaggctagtggtcaacatcacagagactcagcccccaacaaacttctggctg 1274
DB 1012 ggtgaatgtagcgttctacaccccttgagctgaggaactcttccaaactcttgaggcat 1071
QY 1275 catcgaagggcgcgtcagagccaagactacgtgtgctatcggggccaagagggagtgat 1334
DB 1072 catccggggggcgtgtagagcctgacgtacgctgagtgatgaggaacacagcagacactg 1331
QY 1335 gggcccaaggaagcctaatacgcgtgtgggagcgtatactcctgagactgtgtggagc 1394
DB 1132 ggtgacaggggctgtgagaccacagacagtgccacggcgttcttgagcttccctgctg 1191
QY 1395 cttctcctcagtgtagcaagc-----cttcggcccgagagaagctcctcttcacag 1451
DB 1192 cctggggacccctctgaagaaaggacactggtccttccgagatcaactctgtttgcgag 1251
QY 1452 ctggaagcgtgtgagcttggaaagcgtgctgctcacaagagtggtcgaagcctacacag 1511
DB 1252 ctggggggcgtgagagagtttggtcctatggcttcacacagatttcacagaaagattcttca 1311
QY 1512 cgtgtcgcacccaagacgtagtgtagcgtgagcctggaacagcagtgctgaggagatga 1571
DB 1312 caagcttcagagacgcgcagcgttggtcctacatcaacgtggagacatcgttggttccaaagc 1371
QY 1572 caagtttcagtcagaagaccgcccctcttgacaagtcttatgtgagagtgctctgaagca 1631
DB 1372 tacctttaggggtcagaggagcgcctccctgtccagagcgtgcttcttcaaaccaaga 1431
QY 1632 ggtggaattcccaacacacagtgtagagactcctatgaacaggtgtgttcaacaaatcc 1691
DB 1432 gatccgctacacagaccccttgccgacacacgacatctacagaaactgacatccggaacttca 1491
QY 1692 cagctggatgtcgtgagtgatccggccctaccacatga-----cagcag 1736
DB 1492 cccgacagaccccggtgacccctggtgcccacacacttgggttctgtgagctctggcagacga 1551
QY 1737 tgcctatcctcaagccttctgtgagagtcctgtccgttgagttcctccttatggagga 1796
DB 1552 ctatgacaccttgccttcttccctggacatctcctcatggaattgcttactatgata 1611
QY 1797 c-----gaccaggtctaacatcctcctgacacaaagagaaactta 1841
DB 1612 ccggagcaaacatttcagccacagatctacccacttaccacacagccttcttgcacactttga 1671

QY 1842 gaacctgataaagtgtcgtcaagcgccgctgcccgcgtggtgcccagcgccgtgaccagt 1901
 DB 1672 CTATGTGACACAGTTTGTGGACCCGGGCTT---CAGCACCATCAGGCTGTGGCCGGGAC 1728
 QY 1902 cgcagggcagctccctccatcagcgtcagccagatcgctgtgcccctgacttcgccc 1961
 DB 1729 AGCGGGAGAGTGTGATCTCCGGCTCAGTGACAGCTTCTTCGCCCCCAAGTCAGTGA 1788
 QY 1962 ctacgggggagcgtcgtctcagagacatcgggaaacctcaagagttctctgggagctcaa 2021
 DB 1789 CTACAGTGAGACACTCCGAGCTTCCTCGAGGAGCCGACAGATCTTGGGGCCTGCT 1848
 QY 2022 gg 2023
 DB 1849 GG 1850

DB 516 AGANNCGTA 525

Search completed: January 13, 2000, 02:55:32
 Job time: 2290 sec

RESULT 15
 LOCUS 123810 123810 660 bp DNA PAT 07-OCT-1996
 DEFINITION Sequence 28 from patent US 5538866.
 ACCESSION 123810
 NID 91603680
 VERSION 123810.1 GI:1603680
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 660)
 AUTHORS Israelli,R.S, Heston,W.D.W. and Fair,W.R.
 TITLE Prostate-specific membrane antigen
 JOURNAL Patent: US 5538866-A 28 23-JUL-1996;
 FEATURES
 source 1.660 /organism="unknown"
 BASE COUNT 190 a 126 c 147 g 181 t 16 others
 ORIGIN

Query Match 3.1%; Score 88.8; DB 5; Length 660;
 Best Local Similarity 56.0%; Pred.No.1.4e-05;
 Matches 241; Conservative 0; Mismatches 180; Indels 9; Gaps 4;

QY 1217 tgcgctagtggtcaacaatcacagacctccaccccatcaacaacatcttcggctgca 1276
 DB 96 TGAAGCTCACTGTGAACATGTACTGAAGAACAAGAAATACTTAACATCTTTGGCGTTA 155
 QY 1277 tgcgaagcgccgtcagagccagatcacatcgtgtcactcgtgggcccagagagatgcatgg 1336
 DB 156 TTAAGGCTATGAGGACACGCGCTACATGTAGTAGAGGCCAGAGAGCGTTGGG 215
 QY 1337 gcccaagg-agcagctaaatccgctgtgggagc-ggctatactcgtgagcgtgtgcgagc 1394
 DB 216 GCCCTGGTNGTCCGAAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 275
 QY 1395 ctttcctccatggtgagcaac---ggcttcgggcccgcagagatctccttcaatcag 1451
 DB 276 ATTCTCAGATATGATTTCAAAAGATGATTTAGACCCACAGAGATATATCTTTGCCAG 335
 QY 1452 ctggagcaggtgggtgacttggagcgtgggctccacgagatgtagaagctaacctcag 1511
 DB 336 CTGAGCTGACAGAGACTATGAGCTGTGTCCGACTGAGTGAGGGGTACTTTC 395
 QY 1512 cgtgtcgaacctcaagccg----tagtgtacgtgagcctgagacaacgagtcgtgggg 1567
 DB 396 ATCTTGCATCTAAAGNNNGCTTTCATACATTAATNCTGATAAAGTCGTCGGGTA 455
 QY 1568 atgacaagttcatgccaaagcccccctctgacaagttcatltgagagtgctcctga 1627
 DB 456 CTAGCAACTTCAAGGTTTCTGCCAGCCCTTATATATACACTTATGGGAAAGATATGC 515
 QY 1628 agcaggtgga 1637

Fri Jan 14 12:05:06 2000

us-09-358-755-2.rge

This Page Blank (usp10)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 13, 2000, 18:56:04 : Search time 120.47 Seconds

(without alignments)
5974.953 Million cell updates/sec

Title: US-09-358-755-2

Perfect score: 2877
Sequence: 1 ctgcagcgtctcagaggggga.....ctgatcataatgaaccacaa 2877

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|---------------------|
| 1 | 148.2 | 5.2 | 779 | 1 | Q65535 | Sequence used in 1 |
| 2 | 127 | 4.4 | 540 | 1 | Q65537 | Sequence used in 1 |
| 3 | 89.4 | 3.1 | 93 | 1 | T19421 | Human gene signatu |
| 4 | 88.8 | 3.1 | 660 | 1 | Q65536 | Sequence used in 1 |
| 5 | 53.8 | 1.9 | 2653 | 1 | Q65520 | Prostate-specific |
| 6 | 53.8 | 1.9 | 2654 | 1 | T36785 | Prostate-specific |
| 7 | 46.2 | 1.6 | 1030 | 1 | V99230 | DNA encoding an ac |
| 8 | 46 | 1.6 | 49272 | 1 | V35000 | Mycobacteriophage |
| 9 | 45.8 | 1.6 | 53789 | 1 | V21187 | Amycolatopsis medi |
| 10 | 45 | 1.6 | 4356 | 1 | Q37543 | Cardiac adenyllyl c |
| 11 | 45 | 1.6 | 4356 | 1 | Q95540 | Cardiac adenyllyl c |
| 12 | 44.8 | 1.6 | 5821 | 1 | T58686 | DNA encoding S. fr |
| 13 | 44.8 | 1.6 | 2135 | 1 | X26879 | DNA encoding a pro |
| 14 | 44.6 | 1.6 | 1164 | 1 | V32494 | S. griseofuscus gl |
| 15 | 44.4 | 1.5 | 2638 | 1 | V62134 | HSV-2 strain SB5 C |
| 16 | 44.4 | 1.5 | 20387 | 1 | V62159 | HSV-2 strain SB5 C |
| 17 | 44.4 | 1.5 | 11723 | 1 | V62176 | Human adenovirus A1 |
| 18 | 44.4 | 1.5 | 114955 | 1 | X53491 | Chimeric herregulin |
| 19 | 44 | 1.5 | 1389 | 1 | T18534 | Amycolatopsis medi |
| 20 | 44 | 1.5 | 5676 | 1 | V21186 | Amycolatopsis medi |
| 21 | 44 | 1.5 | 53789 | 1 | V21187 | Chlamydomonas rein |
| 22 | 43.8 | 1.5 | 2846 | 1 | V41260 | Platanolide syntha |
| 23 | 43.2 | 1.5 | 44377 | 1 | T78508 | Bleomycin acetyl t |
| 24 | 43.2 | 1.5 | 44377 | 1 | T80414 | S. clavuligerus cl |
| 25 | 43 | 1.5 | 1278 | 1 | Q21629 | Manic-depressive i |
| 26 | 42.8 | 1.5 | 15079 | 1 | Q91580 | S. erythraea oleand |
| 27 | 42.8 | 1.5 | 8065 | 1 | V38335 | Human Sox1 encodin |
| 28 | 42.8 | 1.5 | 6093 | 1 | X25775 | DNA fragment conta |
| 29 | 42.6 | 1.5 | 1542 | 1 | X16151 | ACY B2 gene - enco |
| 30 | 42 | 1.5 | 2729 | 1 | N92629 | Infected cell prot |
| 31 | 42 | 1.5 | 1164 | 1 | N92631 | The nucleotide seq |
| 32 | 42 | 1.5 | 1630 | 1 | Q66684 | Streptomyces venez |
| 33 | 42 | 1.5 | 12001 | 1 | Q76213 | Elastase gene. Pse |
| 34 | 42 | 1.5 | 4257 | 1 | V10362 | Sequence encoding |
| 35 | 42 | 1.5 | 4257 | 1 | V68520 | Human K+ channel 2 |
| 36 | 41.6 | 1.4 | 15872 | 1 | Q04529 | |
| 37 | 41.6 | 1.4 | 2072 | 1 | Q04529 | |
| 38 | 41.6 | 1.4 | 3059 | 1 | N70566 | |
| 39 | 41.6 | 1.4 | 2483 | 1 | T12462 | |

| | | | | | | |
|----|------|-----|-------|---|--------|--------------------|
| 40 | 41.6 | 1.4 | 2483 | 1 | V04874 | CDNA sequence of t |
| 41 | 41.6 | 1.4 | 1831 | 1 | X26880 | DNA encoding a pro |
| 42 | 41 | 1.4 | 1564 | 1 | Q11415 | Ryanodine receptor |
| 43 | 41 | 1.4 | 1960 | 1 | T43380 | Human cytokine res |
| 44 | 41 | 1.4 | 43280 | 1 | T80413 | Tyrosine synthase |
| 45 | 41 | 1.4 | 17955 | 1 | V56642 | Actinoplanes sp. a |

ALIGNMENTS

RESULT 1
065535
ID 065535 standard; cDNA: 779 BP.
AC 065535;
AT 11-JAN-1995 (first entry)
DE Sequence used in identification of PSM Ag cDNA.
KW Prostate-specific membrane antigen; PSM: prostate cancer;
KW transmembrane glycoprotein; imaging; targeting; tumour detection;
KW antibody detection; ds.
OS Homo sapiens.
PN W09409820-A.
PD 11-MAY-1994.
PF 05-NOV-1993: U10624.
PR 05-NOV-1992: US-973337.
PA (SLOK) SLOAN KETTERING INST CANCER.
PI Fair WR, Heston MDW, Israeli RS;
DR WPI: 94-167129/20.
PT Prostate-specific membrane antigen and DNA encoding it - is
PT useful for detecting haematogenous micro-metastatic tumour cells
PT and for identifying ligands which bind to PSM Ag
PS Disclosure: Page 119-120, 196pp: English
CC This sequence is given in the specification but is not mentioned in
CC the disclosure. The PSM coding sequence is useful for suppressing or
CC modulating the metastatic ability of prostate tumour cells to grow,
CC or for eliminating them. The protein is useful to identify or purify
CC ligands of the Ag. It is also an attractive target for Ab-directed
CC imaging and targeting of prostatic tumour deposits. 175 T;
SQ Sequence 779 BP: 197 A; 182 C; 201 G; 175 T;

Query Match 5.2%; Score 148.2; DB 1; Length 779;
Best local similarity 56.9%; Pred. No. 7.6e-22;
Matches 390; Conservative 0; Mismatches 265; Indels 30; Gaps 6;

| | | | |
|----|------|---------------------------------------------------------------|------|
| QY | 988 | ggaacatgtcacccttgggaacatggagacccctacacacactgtctccctcctcaatcaa | 1047 |
| DB | 17 | GGACATGCCACCTTGGAACTGAGACCCCTTACACCCAGGC-TCCCTTGTCAACAC | 75 |
| QY | 1048 | accaca---gtccctccagttgcataatcagccctcccaagatcccaagccac | 1104 |
| DB | 76 | ACCCANNNGTTTCCACCATGTAATCTTCAGGACTACCCACATGCTGTTCAGACATC | 135 |
| QY | 1105 | agagcagacatctccctccgctcgtcgaggaagctcaag--gcctgtgcccccaag | 1162 |
| DB | 136 | TCTAGCATGTCACAGCAGCAGCTGTTCAAGCAATGAGTGAACATGCTTGANAGN | 195 |
| QY | 1163 | aatggcagggagcctcctcctcagctccctatacactg-----gac | 1203 |
| DB | 196 | NGTTGGAAGGTCGATCCANNNTTCTCTAGCTTNGACANNNAACAAGCAGAGANNNN | 255 |
| QY | 1204 | cccgagccagactcggcgtagtgtaacaatcaagagcctccaccatcaaac | 1263 |
| DB | 256 | GCCAGANTTAATGCTGAACATGATGTGAACAATTCATGAAGAGAGATTCTGAC | 315 |
| QY | 1264 | atttggcgtcgtcgtgaagccgctcagagccagatcactacgtgtcaggggcccag | 1323 |
| DB | 316 | ATCTTGGTGTCTATCCAGGAGATTGAAGAACCTGATGGTATGTGATGGAGCCAG | 375 |
| QY | 1324 | aggatcagtgggcccaagagcctaaatcccgctgtggtggcagcgtacatcctcag | 1383 |
| DB | 376 | AGAGACTCCTGGGGCCAGAGAGTGGCTTAAGCTGGACCTGTAATATTGTTGGA | 435 |

OY 1384 ctggtcgagacattctccatggtgaagcaac---ggcttcggcccccagaaatctc 1440
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 436 CTGGCCCGGTATCTCAACATAGTGAAGAAACGAGGGCTTAACAAACCGGCGAAGCATC 495
 OY 1441 cttctcatcagctggagcgtgtgactttgaaagcgttggtccacggagtgctagaa 1500
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 496 AACTTTGTAGCTGAGAGTCAAGAGACTACGAGCTGTGGGTGTACTAATAGCTGGAG 555
 OY 1501 gctctccatcagctgtctgacactcaaaagcgtagtgtatgta--gcctggacaacgag 1558
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 556 GGGTACTGTGCATGCTGCATGCCAAAGCTTTCACCTACATCANNCGTATGCTCCAG 615
 OY 1559 tgcctggggatgacaaagtctcatgccaagcagcccccttcgacaagtctcatgaga 1618
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 616 TCCTGGAGCAAGCCATGTCAGATTTCTGCCAGCCCTTGTGTATATGCTGTGGGA 675
 OY 1619 gtgtcctgaagcaggtggaattctc 1643
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 676 GATATATGAAGGGGTGAAGATCC 700

RESULT 2

065537
 ID 065537 standard; cDNA: 540 BP.
 AC 065537;
 DT 11-JAN-1995 (first entry)
 DE Sequence used in identification of PSM Ag cDNA.
 KW Prostate-specific membrane antigen; PSM: prostate cancer;
 KW transmembrane glycoprotein; imaging; targeting; tumour detection;
 KW antibody detection; ds.
 OS Homo sapiens.
 PN WO9409820-A.
 PD 11-MAY-1994.
 PF 05-NOV-1993; U10624.
 PR 05-NOV-1992; US-973337.
 PA (SLOK) SLOAN KETTERING INST CANCER.
 PI Fair WR, Heston MDW, Israeli RS;
 DR WPI: 94-167129/20.
 PT Prostate-specific membrane antigen and DNA encoding it - is
 useful for detecting haematogenous micro-metastatic tumour cells
 and for identifying ligands which bind to PSM Ag
 PS Disclosure: Page 121; 196pp: English
 CC This sequence is given in the specification but is not mentioned in
 CC the disclosure. The PSM coding sequence is useful for suppressing or
 CC modulating the metastatic ability of prostate tumour cells to grow,
 CC or for eliminating them. The protein is useful to identify or purify
 CC ligands of the Ag. It is also an attractive target for Ab-directed
 CC imaging and targeting of prostatic tumour deposits.
 SQ Sequence 540 BP; 152 A; 106 C; 133 G; 147 T;

Query Match 4.4%; Score 127; DB 1; Length 540;
 Best Local Similarity 60.4%; Pred. No. 1.5e-17;

Matches 261; Conservative 0; Mismatches 166; Indels 5; Gaps 3;

OY 1217 tgcgctagtgtcaacaacacagactccaccccccaacaacattcgtgtga 1276
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 75 TGAAGCTACTGTGACCAATGTGCTGAAGAGATTAATAATTTTAAACATCTTGGAGTTA 134
 OY 1277 tcaagggccgctcagagcagatacactacgtgtcactcggggcccaagagatgcaagg 1336
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 133 TTAAGGCTTTGTAGAACCAAGATCACTATGTGTAGTTGGGCCCCAGAGATGATGGG 194
 OY 1337 gcccaagagcaagtaaat-cgctgtggggacggtatactcctcgtgagctgtgcggacc 1395
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 195 GCCCTGGACCTCAAAATNCGCTGTAGGCACAGCTCTCTATTGAACCTGCCAGATG 254
 OY 1396 ttcttcctcatgg---ttagcaagcgttcggcccgcaagagctcctctcctcatgcgc 1452
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 255 TTCTAGATATATGTTTAAAGATGGGTTTCAGCCAGCAAGACATTAATCTTTGCCAGT 314
 OY 1453 tggagcgtgtggtacttgaagcgtgggtccacgagatggtcgaaggtacc-tcag 1511
 || || || || || || || || || || || || || || || || || || || || || ||

DB 315 TGAAGTGTGAGAGACTTTGATGCGTTGTGTGCACTGAATGGCTAGAGGATTAACCTTCG 374
 OY 1512 cgtgtcgaacctcaaaagccgtagtgcagtgagcctgagacaacgacgtcgtgggagtga 1571
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 375 TCNCTGATTTAAAGGCTTTCACCTATATTAACCTGATTAAGGCGTCTTGTGACAG 434
 OY 1572 caagttcatcgaagcagccccctctcgaacgtctcattgagatgtcctgaagca 1631
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 435 CAACCTCAAGGTTTCTGTCGACGCCACTGTGTATACGCTTATGGAAGAAACATGCCAAA 494
 OY 1632 gttggaattctc 1643
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 495 TGTGAAGCATCC 506

RESULT 3

119421
 ID T19421 standard; cDNA to mRNA; 93 BP.
 AC T19421;
 DT 04-JUL-1996 (first entry)

DE Human gene signature HUMGS00446.
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.

OS Homo sapiens.
 PN WO9514772-A1.
 PD 01-JUN-1995.
 PF 11-NOV-1994; J01916.
 PR 12-NOV-1993; JP-355504.
 PA (MATS/) MATSUBARA K.
 PI (OKUB/) OKUBO K.
 PT Matsubara K; Okubo K;
 DR WPI: 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function. By preparing a library that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues

PS Claim 1; Page 380; 2245pp: Japanese.
 CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 783 "GS" sequences
 CC given in T19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (gene signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

SQ Sequence 93 BP; 28 A; 17 C; 24 G; 23 T;

Query Match 3.1%; Score 89.4; DB 1; Length 93;
 Best Local Similarity 97.8%; Pred. No. 5.1e-10;
 Matches 90; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2779 gatcgtggcaccatagcctatgagcacaagcgtgtcgtggaagggcgtgagt 2838
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 1 GATCGTTGGCACCATAGGCTTATGGCCAAAGGTGTGTGTGTAAGGGGCGGTGAGT 60
 OY 2839 ttcaatatcaataaaccacccatgatacaataa 2870
 || || || || || || || || || || || || || || || || || || || || || ||
 DB 61 TTCAATATCAATFAAACCAACCTGATATCAATAA 92
 RESULT 4
 ID 065536 standard; cDNA: 660 BP.
 AC 065536;

23-FEB-1996: U02424.
 PR 24-FEB-1995: US-394152.
 PR 02-JUN-1995: US-470735.
 PR 02-JUN-1995: US-466381.
 PA (SLOK) SLOAN KETTERING INST CANCER RES.
 PI Fair WR, Heston MDW, Israeli RS;
 DR WPI: 96-402365/40.
 P-PSDB: W02234.
 CC DNA encoding alternatively spliced prostate-specific membrane
 PT antigen - useful to develop prods. for detecting haemotogenous
 CC micrometastatic tumour cells, or prostate cancer progression
 CC Example 1: Fig 47A-D: 284pp; English.
 CC A cDNA clone (T36785) codes for human 100 kDa prostate-specific
 CC membrane (PSM) antigen (W02234), an integral membrane glycoprotein
 CC that is very highly expressed in prostatic tumours and metastases.
 CC It was obtcd. from lymph node carcinoma of prostate (LNCaP) cell
 CC mRNA by PCR amplification (see also T36795-808) and screening of an
 CC LNCaP cDNA library using an amplified cDNA partial clone as probe.
 CC The cDNA can be used to provide probes and primers useful e.g. in
 CC detecting haemotogenous micrometastatic tumour cells and determining
 CC prostate cancer progression (see also T36827-30 and T36813-18),
 CC and in gene therapy. An alternatively spliced PSM, PSM', has a
 CC shorter cDNA sequence. PSM genomic DNA is given in T36786.
 SQ Sequence 2654 BP: 782 A; 525 C; 639 G; 708 T;

Query Match 1.9%; Score 53.8; DB 1; Length 2654;
 Best Local Similarity 48.8%; Pred. No. 0.016;
 Matches 145; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 1255 atcaacaacatctcggtcgatcgagagccgtcagagccagatacactcgtgtcattc 1314
 DB 1324 ATTTAAATGATGATAGTACTCTCAGAGAGCGAGTGAACACAGATATGATTCG 1383
 QY 1315 gggggccagagagatgcatggtggccagagcagctaaatccgctgtggggagcgctata 1374
 DB 1384 GGAGGTACCGGAGCTCAGTGGTGTGGTGTATGACCCCTCAGATGTGAGAGCGTGT 1443
 QY 1375 ctctctgagctgtgtcggtcgttctccatgtgtgagcaagcgctccggcccgaga 1434
 DB 1444 GTTACGAAATGTTGAGGAGCTTTGGAACACTGAAAAAGAGAGGAGGAGACTAAGA 1503
 QY 1435 agctccctcttcacgcgtggagcggtgtgactttggaagcggtgggtcccaagagtg 1494
 DB 1504 ACAATTTGTTGTCAGAGCTGGGATGCAGAAAGATTGCTTCTTGTCTTACTGTAG 1563
 QY 1495 ctagaaggtcactcagcgtgtcgtcactcaagccgtagtgacgtgagcctggac 1551
 DB 1564 GCAGAGGAGAAATTCAGAGCTCTTCAAGAGCGTGGCTTATATATGCTGAC 1620

RESULT 7
 V99230
 ID V99230 standard; DNA: 1030 BP.
 AC V99230;
 DI 09-MAR-1999 (first entry)
 DE DNA encoding an active acyltransferase domain (venAT) from a PKS.
 DI Enzymatically active acyltransferase domain; polyketide synthase; PKS;
 KW polyketide synthetis; antimicrobial; anticancer; antifungal;
 KW immunosuppressant; antihelminthic; venat; malonate acyltransferase domain;
 KW PKS cluster; Streptomyces venezuelae ATCC15439; ds.
 OS Streptomyces venezuelae.
 PN W0951695-R2.
 PD 19-NOV-1998.
 PF 13-MAY-1998: U09518.
 PR 16-MAY-1997: US-858003.
 PA (ABBO) ABBOTT LAB.
 PI Kakevas SJ, Katz L, Pereda-Lopez A, Ruan X, Staasi DL,
 PI Summers RG;
 DR WPI: 99-070114/06.
 P-PSDB: W87713.
 PT New polyketides produced by microorganism having altered polyketide
 PT synthase genes - especially chimeras containing a heterologous

PT acyltransferase domain, useful as e.g. antimicrobials, anticancer
 PT agents
 PS Claim 24; Fig 13; 137pp; English.
 CC The present sequence encodes an enzymatically active acyltransferase
 CC domain, venAT, from a polyketide synthase (PKS). The venAT protein
 CC is derived from the malonate acyltransferase domain from the
 CC PKS cluster of streptomyces venezuelae ATCC15439. The sequence
 CC can be used to replace acyltransferase domains of other PKS enzymes,
 CC which are then used to generate novel polyketides by de novo biosynthesis
 CC rather than by chemical modification. Polyketides of the invention have
 CC the methyl groups on the macrolactone ring substituted with -H, -Et
 CC and/or -OH, and the ethyl side chain substituted with hydroxymethyl or
 CC dihydroxycyclohexylmethyl side chains. Polyketides of the invention are
 CC antimicrobial, anticancer, antifungal, immunosuppressant and/or
 CC antihelminthic agents, or their intermediates.
 SQ Sequence 1030 BP: 132 A; 383 C; 362 G; 153 T;

Query Match 1.6%; Score 46.2; DB 1; Length 1030;
 Best Local Similarity 47.4%; Pred. No. 0.48;
 Matches 138; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 694 cagctgcgcgtggagaccctgactctactgcccctacagcgccatcggaacgtcacg 753
 DB 238 CMTTCGGCTGTTGAGGCCCGAGGCTCTGCCGACTACTCTATGCGCACCTCATTC 297
 QY 754 ggaagagctgtgtacgcccactaagggccggaagactgcaagactcgtggccagg 813
 DB 298 GCGGAGATGACCCGCGCCGACCTGCGCGGGCTCTCGATCTGGCGAGCGCGTCTCTG 357
 QY 814 ggcgtgagatccagctgtggccgctgtgctgtgctgtgctgtgctgtgctgtgctgt 873
 DB 358 GTGCGCCACCGCGCGCGCTGATGCACTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 417
 QY 874 aaggtgaccatgctcagagactcgggtcgaagagtgctacataaccagagcagcg 933
 DB 418 CAGCGGAGGAGAGAGAGGTACGAGCGCGCTGCGGACTTGCAGAGATGGCGTTCG 477
 QY 934 gactctcccaagaccaccaagccaagcctgtctcagcagcagcagcg 984
 DB 478 GCGGAGATCAGACGCCCGAGCCGACCTGCTCTCGCGGAGAGAGCGCG 528

RESULT 8
 V35000
 ID V35000 standard; DNA: 49272 BP.
 AC V35000;
 DI 27-AUG-1998 (first entry)
 DE Mycobacteriophage D29 DNA.
 DI Shuttle plasmid; vector; mycobacteriophage; lambda phage cosmid; antigen;
 KW anti-tumour agent; lymphokine; immune response; tolerance; mutant; toxin;
 KW autoimmune disease; antiserum; vaccine; humoral response; T-cell memory;
 KW cellular response; mucosal response; diagnostic; antibody;
 KW infectious agent; ds.
 OS Mycobacteriophage D29.
 PN US5773267-A.
 PD 30-JUN-1998.
 PF 07-MAR-1996: 614770.
 PR 07-MAR-1996: US-614770.
 PR 29-FEB-1992: US-833431.
 PR 29-APR-1993: US-057531.
 PR 23-MAY-1994: US-247901.
 PA (UPTI) UNIV PITTSBURGH.
 PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 PI Hatfull GF, Jacobs WR;
 DR WPI: 98-387007/33.
 PT Nucleic acid encoding the D29 mycobacteriophage and related shuttle
 PT plasmids - for producing recombinant mycobacteria and auxotrophic
 PT cell memory.
 PS Claim 1; Column 7-50: 28pp; English.
 CC This sequence encodes the D29 mycobacteriophage which is used in the
 CC construction of a new shuttle plasmid containing D29 with an E. coli

| | |
|----------------------------------------------------------------|------------------------------------------------------------------------------|
| PA | (AMCY) AMERICAN CYANAMID CO. |
| DR | WPI: 95-214006/28. |
| DR | P-PSDB; R78519. |
| PT | Cardiac adenylyl cyclase and corresp. DNA - having specified sequences |
| PS | Claim 1; Fig 2; 45pp; Chinese. |
| CC | Q93540 encodes R78519, the novel effector enzyme cardiac adenylyl cyclase. |
| SC | Sequence 4356 BP; 836 A; 1372 C; 1373 G; 775 T; |
| Query Match 1.6%; Score 45; DB 1; Length 4356; | |
| Best Local Similarity 49.4%; Pred.No.1.1; | |
| Matches 117; Conservative 0; Mismatches 120; Indels 0; Gaps 0; | |
| QY | 1864 ggcgcccctcccgccggctggcacaagcgttgcccccaagctcgaggcgacatctcatccgg 1923 |
| Db | 469 gcccttcgggaccttgggctgcgcgcgcgcccccggcccttcgtccaagccactccac 410 |
| QY | 1924 ctcaagcacagaatcgactgtctgcctccctgaattcgcgctaaagggagatcgtctccag 1983 |
| Db | 409 cgaagcggggggcgaccatctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcag 350 |
| QY | 1994 cacatcgggaaccttaacgatlctctggygaacctcaaagcccgcyggtcgaacctcgag 2043 |
| Db | 349 ggccgccccggagatcccccccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 290 |
| QY | 2044 tgggtgactctcgcgcgcggggagactacatactcgcgcgcggaagaagtctcgcgagag 2100 |
| Db | 289 caGTCGTGCCTGGCGGCTCCTGCCAGGCCACCTGTGAAGCGAAGTAAAGCCGAGG 233 |
| RESULT 12 | |
| ID | T58686/c |
| DC | T58686 standard; DNA: 5821 BP. |
| DT | 17-FEB-1998 (first entry) |
| DE | DNA encoding S. fradiae tylosin biosynthesis gene products. |
| KM | Tylosin: biosynthetic gene product; production; antibiotic; ds. |
| OS | Streptomyces fradiae. |
| FH | Key Location/Qualifiers |
| FT | CDS complement(73..1260) |
| ET | /tag= a |
| FT | /product= tylE |
| FT | complement(1277..1957) |
| FT | /tag= b |
| FT | /product= tylD |
| FT | complement(2286..4049) |
| FT | /tag= c |
| FT | /product= tylH |
| FT | 4094..4927 |
| FT | /tag= d |
| FT | /product= tylF |
| FT | complement(5046..5660) |
| FT | /tag= e |
| FT | /product= tylJ |
| PN | US5672497-A. |
| PD | 30-SEP-1997. |
| PF | 21-DEC-1985; 575843. |
| PR | 12-MAY-1989; US-351350. |
| PR | 21-MAR-1986; US-842330. |
| PR | 25-JUL-1986; US-890670. |
| PR | 24-FEB-1987; US-018237. |
| PR | 06-AUG-1981; US-744222. |
| PR | 28-JUL-1993; US-107232. |
| PR | 17-FEB-1994; US-198672. |
| PR | 21-DEC-1995; US-575843. |
| PA | (ELIL) ILILY & CO ELI. |
| PI | Cox KL, Fishman SE, Hersberger CL, Seno ET; |
| DR | WPI: 97-48860/45. |
| DR | P-PSDB; W33272, W33273, W33274, W33275, W33276. |
| PT | DNA encoding Streptomyces fradiae tylosin biosynthesis gene products |
| PT | - for increasing tylosin production in Streptomyces spp. |

| | | | |
|------------------|--------------------------------------------------------------|------------------------------------------------------------------------|-------|
| OY | 370 | gttgtgtgctcagtcagtgaagatcgtcaactatbagcccttagaccgtgatcttccaaccaaggacaga | 429 |
| Db | 460 | CTCGACCCGCATGAAAGAGAGGCCTTTGACTCCTTCGGCGAATGACGTACACCGCCCAAGSGCTAC | 51.9 |
| OY | 430 | ctctaacttgaaagcgacctcacaagccatgtttcttcgtcaattcctctggggaggggcgacctgag | 489 |
| Db | 520 | GACCCTCCCTTTCGGCATGTGACAGCCCCAAGGCCAACAGACGCCCGCGCGCATCTCTCTCCC | 57.9 |
| OY | 490 | gacaccatcagcagaacacagccttcggaaagggtgtgcaggtcgtgcgcggatgacgtgcgt | 54.9 |
| Db | 580 | ACCGTCGGCCACAGCCCTTGACTTATCGAGAGCCTCGAGCGCCCGGAGAGCTGTACGGCGTC | 63.9 |
| OY | 550 | ctgactcagaagaatctgcgcgcgcctctcccgcagaagaagtgtgacacacgtgtgagccgac | 60.9 |
| Db | 640 | AACCGGAGGTCTCGCGCACAGACAGATGGCCGGCTGMAACTTCCCGCANCGGATGGCGAG | 69.9 |
| OY | 610 | acgcaactacgtggtggtctgtcaattccccgatalcccgctcaaccccaaacacctgcaactgtgc | 66.8 |
| Db | 700 | GCCCTGTGGCGGGCGAAGCTTTCTCCATCGACTGACTCAACGGCCGCTCCGGCATTAAGTAC | 75.9 |
| OY | 670 | gatggtgcgcgggaaagtgtcgtgagagacagctgtgcgtcttgagagacccctgacgttactgcc | 72.9 |
| Db | 760 | GACCGAGACCTCGGTTTGGCGCGCGGACACTCGCGGGCGGGTCTGTGGCTGTGTGACTTC | 81.9 |
| OY | 730 | tacagcgcacatcgtgcgaacgtltaacgtgagagagctgtgtgtaacgccatacaggtgcgcgaa | 78.9 |
| Db | 820 | CTGGAGAGCCCGCGGTATTACGAGAGCGCCGCGGCACTTCGACTCAAGCCGCGGAGCGAG | 87.9 |
| OY | 790 | gacctgcagagacctgcgcgcgcgcagagggcgtgtgatatccagtgtggcgcgcgtgcgt-----g | 84.3 |
| Db | 880 | GACTTCGACGGCGGTGTGGGCTTCGGCGCAGAGGCTGTGCATGCCCACTACCTGTACTCTCAAG | 93.9 |
| OY | 844 | gtgcgcgttgggggtgatatgaattcgcgcccaagaaggtgacaaatgtctcaggaattcggagct | 90.3 |
| Db | 940 | GAGCGCGCGCGCGCTTCGCGCGCGAGACCCGGAGGTGACGAGAGGCCCTGTGGCGCGCGCGGT | 99.9 |
| OY | 904 | caagagatgcatcatatatacccaagagccagcaggaattctcccaagaccacccc | 95.4 |
| Db | 1000 | CTGGACACGACTGCGCCACGCCAGCGCGGCGAGCGGCTGTGAGGCCCTGTGCTC | 105.0 |
| RESULT_15 | | | |
| ID | V62134/c | V62134 standard; DNA; 26338 BP. | |
| AC | V62134: | | |
| DT | 07-DEC-1998 | (first entry) | |
| DE | HSV-2 strain SBS Contig ID 104 | DNA sequence. | |
| KW | HSV-2 strain SBS; immunological response induction; therapy; | | |
| KW | antiviral identification; viral protein inhibitor; ss. | | |
| OS | Herpes simplex virus type 2. | | |
| FH | Key | Location/Oualifiers | |
| CDS | | complement (256..3021) | |
| FT | | /*tag= a | |
| FT | | /product= "ORF#1 protein" | |
| FT | | /note= "encoded protein shown in W72033" | |
| FT | | complement (2895..3908) | |
| CDS | | /*tag= b | |
| FT | | /product= "ORF#2 protein" | |
| FT | | /note= "encoded protein shown in W72034" | |
| FT | | complement (5263..6745) | |
| CDS | | /*tag= d | |
| FT | | /product= "ORF#4 protein" | |
| FT | | /note= "encoded protein shown in W72036" | |
| FT | | complement (6775..8046) | |
| CDS | | /*tag= e | |
| FT | | /product= "ORF#5 protein" | |
| FT | | /note= "encoded protein shown in W72037" | |
| FT | | complement (8386..9477) | |
| CDS | | | |

| | |
|----------|---------------------------------------------------------------------------|
| FT | /tag= f |
| FT | /product= "ORF#6 protein" |
| FT | /note= "encoded protein shown in W72038" |
| CDS | complement (9721..13911) |
| FT | /tag= g |
| FT | /product= "ORF#7 protein" |
| FT | /note= "encoded protein shown in W72039" |
| CDS | complement (14158..14826) |
| FT | /tag= h |
| FT | /product= "ORF#8 protein" |
| FT | /note= "encoded protein shown in W72040" |
| FT | /tag= i |
| FT | /product= "ORF#9 protein" |
| FT | /note= "encoded protein shown in W72041" |
| CDS | complement (17295..19110) |
| FT | /tag= j |
| FT | /product= "ORF#10 protein" |
| FT | /note= "encoded protein shown in W72042" |
| CDS | complement (19111..20064) |
| FT | /tag= k |
| FT | /product= "ORF#11 protein" |
| FT | /note= "encoded protein shown in W72043" |
| CDS | complement (20149..21279) |
| FT | /tag= l |
| FT | /product= "ORF#12 protein" |
| FT | /note= "encoded protein shown in W72044" |
| CDS | complement (20962..22026) |
| FT | /tag= m |
| FT | /product= "ORF#13 protein" |
| FT | /note= "encoded protein shown in W72045" |
| CDS | complement (22307..23887) |
| FT | /tag= n |
| FT | /product= "ORF#14 protein" |
| FT | /note= "encoded protein shown in W72046" |
| CDS | complement (23778..24065) |
| FT | /tag= o |
| FT | /product= "ORF#15 protein" |
| FT | /note= "encoded protein shown in W72047" |
| CDS | complement (24286..25352) |
| FT | /tag= p |
| FT | /product= "ORF#16 protein" |
| FT | /note= "encoded protein shown in W72048" |
| CDS | complement (25457..26215) |
| FT | /tag= q |
| FT | /product= "ORF#17 protein" |
| FT | /note= "encoded protein shown in W72049" |
| PN | W09820016-AI. |
| PD | 14-MAY-1998. |
| PE | 31-OCT-1997; U20016. |
| PR | 09-JUN-1997; US-049018. |
| PR | 04-JUN-1996; US-030279. |
| PA | (SMIR) SMITHKLINE BEECHAM CORP. |
| P1 | Chan JY, Dabrowski-Amara! CE, Delvecchio AM, Dillon SB, |
| P1 | Esser KM, Leary JT; |
| DR | WPI: 98-286847/25. |
| DR | P-BSDB; W72033, W72034, W72035, W72036, W72037, W72038, W72039, W72040, |
| DR | W72041, W72042, W72043, W72044, W72045, W72046, W72047, W72048, W72049. |
| PT | Hepes simplex virus type-2 sequences - useful in, e.g. prevention |
| PT | and treatment of infection or inducing immunological response in |
| PT | mammal |
| PS | Claim 1: Page 226-237: 748bp: English. |
| CC | This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA |
| CC | sequence of the invention. This sequence was isolated from HSV-2 strain |
| CC | SBS (deposited at ATCC VR-2546), is designated Contig ID 104, and encodes |
| CC | 17 HSV-2 proteins. The proteins can be used for the treatment or |
| CC | prevention of disease, to induce an immunological response in a mammal or |
| CC | to identify inhibitors, activators or novel antivirals. Antagonists of |
| CC | the proteins can be used to inhibit a viral polypeptide. The DNA sequence |
| CC | or a vector containing it can also be used to induce an immunological |
| CC | response in a mammal. |
| Sequence | 26338 BP; 4257 A; 8879 C; 9241 G; 3959 T; |

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 13, 2000, 02:55:32 : Search time 1880.46 seconds

(without alignments)
5727.858 Million cell updates/sec

Title: US-09-358-755-3

Perfect score: 2519
Sequence: 1 gcgcgcgcgcgcgcgcgcctc.....ctgatatacaataagcccaaaa 2519

Scoring table: IDENTITY_NUC

Searched: 780561 seqs, 2137953050 residues

Database: GenEmbl.*

Word size: 0

Number of hits that pass the threshold: 1561122

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl1:*
10: gb_pl2:*
11: gb_pl3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vl:*
17: em_fun:*
18: em_htg:*
19: em_hum1:*
20: em_hum2:*
21: em_in:*
22: em_om:*
23: em_or:*
24: em_ov:*
25: em_pat:*
26: em_ph:*
27: em_pl:*
28: em_ro:*
29: em_sts:*
30: em_sy:*
31: em_un:*
32: em_vl:*
33: gb_htg1:*
34: gb_htg2:*
35: gb_in1:*
36: gb_in2:*
37: em_ba1:*
38: em_ba2:*
39: em_hum3:*
40: em_hum4:*
41: gb_pr4:*
42: gb_htg3:*
43: gb_htg4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|------------|---------------------|
| 1 | 2378 | 94.4 | 2877 | 41 | AF067864 | AF067864 Homo sapi |
| 2 | 713.4 | 28.3 | 227968 | 11 | AF053356 | AF053356 Homo sapi |
| 3 | 550.4 | 21.8 | 580 | 11 | HUMRF50H04 | AF085928 Homo sapi |
| 4 | 303.8 | 12.1 | 2500 | 4 | GGTFR | X55348 G.gallus nr |
| 5 | 279.2 | 11.1 | 5010 | 9 | HSTR | X01060 Human mRNA |
| 6 | 278.6 | 11.1 | 2826 | 10 | HUMTFRR | M1507 Human trans |
| 7 | 254.4 | 10.1 | 2460 | 12 | CRUTRANREC | L19142 Cricetus |
| 8 | 245.4 | 9.7 | 3413 | 12 | RATFRFR | M58040 Rat transfe |
| 9 | 213.4 | 8.5 | 2292 | 12 | MMTRMRNA | X57349 M.musculus |
| 10 | 158.6 | 6.3 | 780 | 5 | I23809 | I23809 Sequence 27 |
| 11 | 146 | 5.8 | 250 | 13 | G14981 | G14981 human STS S |
| 12 | 125.4 | 5.0 | 540 | 5 | I23811 | I23811 Sequence 29 |
| 13 | 103.6 | 4.1 | 2570 | 12 | AF009921 | AF009921 Rattus no |
| 14 | 91.6 | 3.6 | 2320 | 10 | HSAN012371 | AJ012371 Homo sapi |
| 15 | 87.2 | 3.5 | 660 | 5 | I23810 | I23810 Sequence 28 |
| 16 | 73.4 | 2.9 | 3152 | 10 | HSAN012370 | AJ012370 Homo sapi |
| 17 | 67.8 | 2.7 | 1428 | 12 | AF039707 | AF039707 Rattus no |
| 18 | 67.8 | 2.7 | 2348 | 12 | AF040256 | AF040256 Rattus no |
| 19 | 67.8 | 2.7 | 2899 | 12 | RN075973 | U75973 Rattus norv |
| 20 | 66.8 | 2.7 | 2603 | 12 | AF026380 | AF026380 Mus muscu |
| 21 | 57.8 | 2.3 | 1278 | 1 | ABIPDAGEN | X88853 A.brasiliens |
| 22 | 56.6 | 2.2 | 43632 | 1 | SC6A5 | AL049485 Streptomy |
| 23 | 56.2 | 2.2 | 2532 | 3 | AF050502 | AF050502 Sus scrof |
| 24 | 56 | 2.2 | 22449 | 1 | SPSNBCDE | Y15548 S.pristinae |
| 25 | 56 | 2.2 | 22449 | 1 | SPSNBCGEN | X98690 S.pristinae |
| 26 | 55.6 | 2.2 | 306 | 13 | G45215 | G45215 221292.1 ze |
| 27 | 54.8 | 2.2 | 189 | 12 | S81327 | S81327 transferrin |
| 28 | 54.6 | 2.2 | 3041 | 12 | M0STRANS01 | L13969 Mouse delta |
| 29 | 53.8 | 2.1 | 2653 | 5 | I23794 | I23794 Sequence 1 |
| 30 | 53.8 | 2.1 | 2653 | 10 | HUMPSM | M99487 Human prost |
| 31 | 53.8 | 2.1 | 2518 | 41 | AF176574 | AF176574 Homo sapi |
| 32 | 53.6 | 2.1 | 4459 | 1 | TTPEST | X65609 T.thermophi |
| 33 | 53.6 | 2.1 | 7621 | 1 | TTPESTFO | Y15464 Thermus the |
| 34 | 53.6 | 2.1 | 4171 | 1 | TTPESTG | Z12118 T.thermophi |
| 35 | 53 | 2.1 | 38995 | 1 | SCF34 | AL109974 Streptomy |
| 36 | 51.8 | 2.1 | 3508 | 2 | AF069761 | AF069761 Halobacte |
| 37 | 51.8 | 2.1 | 2014 | 2 | AZS1PD | L26240 Azospirillum |
| 38 | 51.8 | 2.1 | 982 | 7 | ZMHSP18K1 | X54075 Maize mRNA |
| 39 | 50.2 | 2.0 | 10295 | 1 | D63799 | D63799 Thermus the |
| 40 | 50.2 | 2.0 | 40745 | 1 | SC151 | AL109848 Streptomy |
| 41 | 49.8 | 2.0 | 24700 | 1 | SCB8 | AL033654 Streptomy |
| 42 | 48.6 | 1.9 | 40105 | 1 | SC5F2A | AL049587 Streptomy |
| 43 | 48.6 | 1.9 | 843 | 7 | ZMHSP18K2 | X54076 Maize mRNA |
| 44 | 48.2 | 1.9 | 4933 | 1 | ABIPDC | X99587 A.brasiliens |
| 45 | 48 | 1.9 | 1800 | 16 | HS1UL3 | M19121 Herpes simp |

ALIGNMENTS

| RESULT | 1 | LOCUS | AF067864 | DEFINITION | Homo sapiens transferrin receptor 2 alpha (TFR2) mRNA, complete cds. | ACCESSION | AF067864 | NID | 95596369 | VERSION | AF067864.1 | SOURCE | GI:5596369 | ORGANISM | human. | REFERENCE | 1 (Dases 1 to 2877) | AUTHORS | Kawabata,H., Yang,R., Hirama,T., Vuong,P.T., Kawano,S., Combatt,A.F. and Koefler,H.P. | TITLE | Molecular cloning of transferrin receptor 2. A new member of the transferrin receptor-like family | JOURNAL | J. Biol. Chem. 274 (30), 20826-20832 (1999) |
|--------|---|-------|----------|------------|----------------------------------------------------------------------|-----------|----------|-----|----------|---------|------------|--------|------------|----------|--------|-----------|---------------------|---------|---------------------------------------------------------------------------------------|-------|---------------------------------------------------------------------------------------------------|---------|---------------------------------------------|
| RESULT | 1 | LOCUS | AF067864 | DEFINITION | Homo sapiens transferrin receptor 2 alpha (TFR2) mRNA, complete cds. | ACCESSION | AF067864 | NID | 95596369 | VERSION | AF067864.1 | SOURCE | GI:5596369 | ORGANISM | human. | REFERENCE | 1 (Dases 1 to 2877) | AUTHORS | Kawabata,H., Yang,R., Hirama,T., Vuong,P.T., Kawano,S., Combatt,A.F. and Koefler,H.P. | TITLE | Molecular cloning of transferrin receptor 2. A new member of the transferrin receptor-like family | JOURNAL | J. Biol. Chem. 274 (30), 20826-20832 (1999) |

```

MEDLINE      99340005
REFERENCE    2 (bases 1 to 2877)
AUTHORS      Kawabata,H., Yang,R., Hiramata,T. and Koeffler,H.P.
TITLE        Direct Submission
JOURNAL      Submitted (21-MAY-1998) Hematology/Oncology, Cedars-Sinai Medical
              Center, 8700 Beverly Blvd, Los Angeles, CA 90048, USA
FEATURES
source
1. 2877
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="7"
   /map="7q22-q31"
   /cell_line="leukemia TF-1"
   1. 2877
   /gene="TFR2"
   28. 2433
   /gene="TFR2"
   /codon_start=1
   /product="transferrin receptor 2 alpha"
   /protein_id="AAD45561.1"
   /db_xref="PID:g5596370"
   /db_xref="GI:5396370"
   /translation="MERLWLFQRAQQLSPRSSQTVYQVHVEPRKGHLEEDDEEG
   AETLAHFCMELRGEPLGRPROPMILPMAAGRAAPYLITALLITGAFILGY
   AFRGSCACGDSLYVSEEDVNEPDLDFQGRLYMDLQAMFLOFEGRLDITROT
   SLRERVASGAGMALFODIRALSRKLDHWTMDTHVGLPPDPHAPMTLHMDVAG
   KYGEOLPLEDPVYCYSAIGNVTGELIYAHYGRPDLODIARCDIPRGLILYRVG
   VISFAKVTNNADFGAOGVLIIPEDPDESODPKPRLSQAQVIGHVILGIDPTPG
   PFSFNQIOPVPAASSGLPIPAOPISADIASRLRLKVPAPQEWQSLSSPHLG
   PEPRLLVVNNHSTPIINNIGIEGRSEPHYVVGAROMGGAASKAVGAIL
   LELVRFSSMVSNGFRRLSLFISMDGDFGSVGTLEWLEGLVYLHKAAYVYSLD
   NAVLGDKRFHAKTSLPLTSLIESVLKOVSPNHSQTLVEOVFTSPMDAVIRLP
   MDSASFTAFVGPVAFERFPMEDQAVPEFIHKETPYNLHKVILGRIPAYAAVNO
   LAGOLLIRSHRLPLDRGRYDVLRIKIGNINEPSGLKRGILLOWLYVARQGYI
   RAABKIROITISSEERDEKLRITNINRIKRIVERFIPLSQVSPDSFRHIEGRDHT
   LGALDHLRLNSNSTGPATVSTGFESRPRRDLALTWTLOGANALSGDVNID
   NNF"
BASE COUNT   571 a 924 c 847 g 535 t
ORIGIN
Query Match 94.4%; Score 2378; DB 41; Length 2877;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2381; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 134 cccgcttagcaaacacccgcttcgggaaaggtggcagcgtcggcggaatggcgctc 133
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 492 CACCATCAGGCAACACGCTTCGGGAACGGGTGGCAGGCTCGCGGCGATGCGCTCT 551
QY 194 gactcaggaattcgcgcgcgctctccgcacagaagctggacacgtgtggacgcagac 253
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 552 GATTCAGGACATTCGCGCGCGCTCTCCGCCGAAGACTGGACACCTGTGGACGACAC 611
QY 254 gcaactcgtgggctgaattccggatccggctcagctcaccccaaacacctgactggtcga 313
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 612 GCACATAGTGGGGCTGCAATTCGCGATCCGGGTACACCAACACCTGCACTGGGGCGA 671
QY 314 tggggccgggaaggtcgggaagcagctgcgctgggaagacctgacgttactatgcccta 373
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 672 TGAAGCGGGGAAGGTGGAAGAGCTGCCGCTGGAGAGACCTGACCTTACTACGCCCTA 731
QY 374 caacgccatcggcaacgtcacgggaagagctggtacgcacactacggggggccgaaga 433
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 732 CAGCGCATGCGCACTCAGGGAAGAGCTGTGTACGCCCACTACGGGGGCCCGAAGA 791
QY 434 cctgcaggaacctgcgggcccagggcgtbga tccagtggcgcgctgctgtggtgcgt 493
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 792 CCTGACAGACCTCGGGGCCAGGGCGGTGATCCAGTGGCGGCTGTGCTGTGGGGCGT 851
QY 494 ggggggatcagcttcgccccagaaggtgaccaatgtccaggaactcggggctcaaggagt 553
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 852 GGGGGTATCATGCTTGCCCGAAGGTGACCAATGCTCAGGACTTCGGGGCTCAAGAGT 911

```

```

QY 554 gtcataataccagagccagcgactctccaggagccaccacaaagccagcgtgtccag 613
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 912 GCTCATATACCCAGAGCCAGCGGACTTCTCCAGAGACCCACCAAGCCTGTGCCAG 971
QY 614 ccagcaggaagtgtatgacatgtgacacttggaaacttggagacccctacacacttgcct 673
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 972 CCAGCAGGAGAGTGTATGACATGTGCACTGGGAAGTGAAGACCCCTACACACTGCTT 1031
QY 674 cccctccctcaataaacccaggtctccctcagttgtacatcaatgaagcctccagcaatccc 733
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1032 CCGTCTCTTCATCAACAAACCAAGTCCCTCCAGTGTGATATACAGGCTTCCAGCATCC 1091
QY 734 agccagcccatcagtgtagacatctgctcccgccgctgtaggaagctcaagggccctgt 793
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1092 AGCCCAAGCCCATCAGTGCAGACATTCCTCCCGCTGCTGAGAGAACTCAAGGCCCTGT 1151
QY 794 ggcgccccaagaatggaaggaggagccctcctaagctcccttcaacttgcacttggccgggac 853
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1152 GGGCCCCCAAGATGGCAGGGAGGCTCCTAGGCTCCCTTATCACTGGCCCGGGGCC 1211
QY 854 aagactcgggctagtgtgtcaacaatcaagaagacctccaccccatcaacacactctgg 913
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1212 ACGACTGCGGCTAGTGTGTCACAAATCACAGACCTCCACCCATCAACACATCTTCGG 1271
QY 914 ctgacatcgaaggccgctcagagccagatcactactgctgtcatcggggcccaaggagatgc 973
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1272 CTGCATCGAAGGCCGCTCAGAGCCAGATACACTGTTGTCACTGGGGCCCAAGGGATTC 1331
QY 974 atgggcccccaaggagcagctaaatccgctgttgggagcgtctatacttcttgaagctgtgc 1033
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1332 ATGGGGCCCAAGAGCAGCTAAATCCGCTGTGGGAGAGGCTATACCTCTGAGACTGGTGG 1391
QY 1034 gaccttctctcagatgtgtgaagaagcgtctccggcccgcaagaatctccttcaatcag 1093
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1392 GACCTTTCTCTCCATGVTGAGCAGCGCTTCGGCCCGCAAGCTCTCTTCATCAG 1451
QY 1094 ctggagcgggtgtgacttgggaagcgttgggtccacaggaatgtgtcagaagctactcag 1153
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1452 CTGGAGCGGTGTGACTTTGGAAGCGTGGGCTCCACGAGTGGCTTAGAAGCTTACTCAG 1511
QY 1154 cgtgtcgcacctcaaaagccgtagtgtacgttgaagcttgcgaacagcagtgcttggggatga 1213
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1512 CGGTGTCGACCTCAAGCCGCTAGTGTAGCTGAGCCTGCAACGCGAGTCTGGGGGATGA 1571
QY 1214 caagtttaatgcgaagaacagcccccctcttgaagaatctcatatggagtgctctgaagca 1273
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1572 CAAGTTTCAATGCCAAMACCAAGCCCTCTTGACAACTCATTTGAGAGTGTCTGAAGCA 1631
QY 1274 ggtggaattcccaaaccaagtgaggagagactctcatgaaagagtggtgttacaatcc 1333
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1632 GGTGGAATTTCCCAACCAAGTGGAGACTCTCTATGAAACAGTGTGTTTCAACCAATCC 1691
QY 1334 cagcttgggaatgtcgaagtgatccgggcccctacccaatggaacagatgtcattctctac 1393
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1692 CAGCTGGATGTGTAGAGTGATCCGGGCCCTTACCAATGAGACAGAGTCTATTCTCTGAC 1751
QY 1394 ggccttcttgggaagtcctcgtccgtcgtgagttctccttataggagagcagcactaccc 1453
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1752 GGCCTTGTGGAGAGTCCCGCTCGAGATTCCTTTAAGGAGGAGACAGACAGGCTTACCC 1811
QY 1454 attcctgcacacaagaaggaagacactatagaaacctgcataagagtgctcagaagccgct 1513
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1812 ATTCTGCAACAAGAGAGACATTATAGAACTGTCAATAGTGTCTGCAAGGCCGCT 1871
QY 1514 gcccgcgttggcccaagccgttggccagcttgcagggcagcctctcaatccggtcagcca 1573
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1872 GCCGCGCTGGCCCAAGCGCTGGCCAGCTCGAAGGCGAGCTCTCATCGGCTCAGGCA 1931
QY 1574 cgatgcctcgtgcgcccctctgacttcggccgctcaagggaagcgtcgtcctcaagcaatcgg 1633
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1932 CGATCGCTGTGTCCTTCTGACTTGGCGGTACAGGGGAGCTGTCTTCAAGCAATCGG 1991
QY 1634 gaacctcaacgagttctcttgggaactcaaggcccgcggtcgtgacctcagtggtgtga 1693

```

| | | | |
|------------|------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|-------------|
| Db | 1992 | GAACCTCAAGAGTTCCTGGGACCTCAAGGCCCGGGGCTGACCTTCGACATGGGTGTA | 2051 |
| Qy | 1694 | ctcgcgccgggggagactacatccgggcgcgcgcgaaaaagctcgcgcaggaatctacagtc | 1753 |
| Db | 2052 | CTCGGCGCGGGGGAGTACATCCGGGGGGGGAAAAAGTTCGGGACGAGAGATCTACAGCTC | 2111 |
| Qy | 1754 | ggagagaaagagagagagactgcacagcatgtcaacacgtgcataatgcygggtgagtc | 1813 |
| Db | 2112 | GGAGAGAAAGAGACGAGGAGACTGCACGCAGATGTCAACAGTGCACATAAATCGGGTGAAGTT | 2171 |
| Qy | 1814 | ctacttccttcccgacgaagtgtccgaagccgcagctcccgcttcgcgcacatcttaatgg | 1873 |
| Db | 2172 | CTACTTCCTTTCCAGTACGTGTGCGCAGCGCACTCCCGTTCCGCGACATCTTATGGG | 2231 |
| Qy | 1874 | ccgtgtagaacacacacgctctgggcgcgcctctgcacacacctcgcgcctgcgtcccaacg | 1933 |
| Db | 2232 | CCGTGGAGACACACACGCTGGGGCGCGCTGCGAGCACACGTGGGGCTGTGCGTCCACAG | 2291 |
| Qy | 1994 | gtcagccctgcctcaaccctgcgcgcgtctgcgaaggagacaaatggcgttaagggagatctg | 2053 |
| Db | 2352 | GCTTAGCCCTGCTCACTTGGACGCTGCAGAGGGGAGCAATGCGCTTAGGGGGAGTGTG | 2411 |
| Qy | 2054 | gaacattgataacaactctcgaagccctcgggagatctccacatcccgcttccccagtaag | 2113 |
| Db | 2412 | GAACATTGATATACAACTTCTGAGGCCCTGGGGATCTCTACATCCTCCGCTCCCAATCAG | 2471 |
| Qy | 2114 | agctccctcgtcctcgtcgttaatgtcaatgcagggtcaggaggttgctcagagtaaccctc | 2173 |
| Db | 2472 | AGCTCCTCTGCTCTCCCTTGAAATGATTCAGGGGTCAAGGAGGTGGCTCAGATCCACCTC | 2531 |
| Qy | 2174 | tcattcgtcgaatcaattctcctaataccctcaacatctccacggagcccgagaccgcgcgc | 2233 |
| Db | 2532 | TCAATTGCTGATCAATTTCTCATTAACCCCTTAACATCTCTCCACGGAGCCCAACCCCAAGC | 2591 |
| Qy | 2234 | acagatatccacacaccccgagccctcgcagtgtagctgcagcccaatgtgaacgtctactg | 2293 |
| Db | 2592 | ACAGATATCCACACACCCCAAGCCCTGAGGTGAGCTGCAGCCCTAATGTAGAGGTCACTG | 2651 |
| Qy | 2294 | tcggttaatcagagtagatgacatccctcaatcaacagccctccctctctcgtgggtctc | 2353 |
| Db | 2652 | TGCGTTAATCAGAGATGATGAGATCCCTTCATATCACACCCCTTCCCTTCTGGGTCTCT | 2711 |
| Qy | 2354 | catatccatagagacacatctctgggaggtttgctgaagccctcggagcctcggcagctcgttaag | 2413 |
| Db | 2712 | CATATCCTTAGAACACCACTCTGGGAGGTTTGTAGGCCCTGGGACCTGGCCAGCTCTGTTAG | 2771 |
| Qy | 2414 | tgaggagagatgcctcgcgcacataagccttaatgcgcacacagtgtagtctgtgtgaaggggc | 2473 |
| Db | 2772 | TGGGAGAGATGCGTCGGCAGCATTACCTTAATGCGCAACAGGTGTGTGTGTAAAGGGGC | 2831 |
| Qy | 2474 | gtggagttcaataataataaacaccctgatalcaataaagccanaa | 2519 |
| Db | 2832 | GTGGAGTTTCAATATCAATTAACCACTGATATCATATAGCCMAAA | 2877 |
| RESULT | 2 | | |
| AF053356 | | | |
| LOCUS | AF053356 | 227968 bp | DNA |
| DEFINITION | Homo sapiens chromosome 7q22 sequence, complete sequence. | | PRI |
| ACCESSION | AF053356 | | 24-NOV-1998 |
| NTD | 93135305 | | |
| VERSION | AF053356.1 | GI:3135305 | |
| KEYWORDS | HTG. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. (bases 1 to 227968) | | |

| | |
|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| AUTHORS | Glockner,G., Scherer,S., Schattewoy,R., Boright,A., Weber,J., Tsai,L.C. and Rosenthal,A. |
| TITLE | Large-scale sequencing of two regions in human chromosome 7q22: analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci reveals 17 genes |
| JOURNAL | Genome Res. 8 (10), 1060-1073 (1998) |
| MEDLINE | 99018118 |
| REFERENCE | 2 (bases 1 to 227968) |
| AUTHORS | Gloekner,G., Rosenthal,A., Drescher,B., Weber,J. and Schattewoy,R. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (05-MAY-1998) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany |
| FEATRES | Location/Qualifiers |
| SOURCE | 1..227968 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7q22" |
| repeat_region | complement(1..197) /note="SINE/Alu, pos. in, rep.: 2..197" /rpt_family="Aluo" |
| misc_feature | 1..169722 /note="PAC clone 3763" /note="Region: 3763" complement(join<309..578,2314..2458,3620..3831, 8788..9180,12095...12222,12821..12985,13079..13326, 14784..14928,15156..15357,16624..16804,18675..18769, 19928..20039,20325..20562,21131..21378,21722..21863, 22884..23036,24183..24361,24472..24582,26030..26173, 27882..27997,28543..28667,29738..29857,30045..30186, 32978..33158,35152..36676,37121..37205,37491..37762, 41053..41140,41322..41593,41773..41919,46462..46608, 46717..46795,47291..47382,48230..>48487)) /product="zonadhesin" complement(309..578) /note="100.000, excellent" |
| exon | complement(join<309..578,2314..2458,3620..3831, 8788..9180,12095...12222,12821..12985,13079..13326, 14784..14928,15156..15357,16624..16804,18675..18769, 19928..20039,20325..20562,21131..21378,21722..21863, 22884..23036,24183..24361,24472..24582,26030..26173, 27882..27997,28543..28667,29738..29857,30045..30186, 32978..33158,35152..36676,37121..37205,37491..37762, 41053..41140,41322..41593,41773..41919,46462..46608, 46717..46795,47291..47382,48230..48383)) /codon_start=1 /product="zonadhesin" /protein_id="AAC78790.1" /db_xref="PID:g3135306" /db_xref="GI:3135306" /translation="MLDPKRNARPGCAVLLSPVLSGGCLSEFHYILRGSGGALL HIYASVLSIRKHTLFSGOPGAPMVAQSVNTYTAARGLOFAVGFGCTPEPAVDAT SIACGGGFPCDPEDNAHPCKMOVSGGCGHAGLCRKNKGCPACGPAAVDAT TQCDFEDAKPLCDMSQVSDDEDMVASGPSPGSTGAAGGIPNGGCSLIHNESNF HRGGVALTFSPDLWEGQLPVHFRAHHNFGISWAQGLLDLILSGEGRVPVLNHWMT QRPMWLTVTPAGFTLPTRLMFEGRGSTAYDILLDLISIRGSCNGHAIYEAE ADEBSHAGOSVRLVSRFCAPADGICVEFAHMGTMGELTGSPADSPPIPKRK RVSGORPYMONTVTVPAGSHGOQMOLIFEGOGSTASYAMGELLINPGCPKVYL ELPPVPSVSSIGSPETGLIENPTISRKPVSLERSVTTERTYKEKPTIPTERK TISTEKPTISEKPNMESEKPTISEKPTITTEKPTIPEKPTISPSEKPTISTEPKPV PTEPTPTPETTYMEPEVIPEKPSIPEKPSIPEKPTISMETIISTEKPTICPP EKPTIPEKPTIPEKPTIPEKSTIEKPTTTEKPTIPEKPTIPEKPTIPEKPTISP KIPEKTIPEKPTIPEKPTISTEEPTPTTEPTIISTEKPSIPEKPTIPEETTYEPT TSVETTISTEKTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTI IEETIISTEKTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEETIISTEKTIPE KPTISPSEKTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKTLARP PSPATGALAVMSPHASPMPMSVILGITTSNSGMSCPPARKSCCAPSACS PRBCGJGLCBEGCVCNCGFLFSIDNHCIQAOSCNCYNNDYEPAKPFNCECHCAC WPGRVCOISDCIGITTHIVCOLKNGOYCAGYAGATLYGDPIYFDGHFFEMKK CTYLAAPCGNSTDPERFRATAKNEOGDEVCISKYVLTLPSTVLTILAGRTLVG QQVTLPIPSKCVPLGASGRFVELQTFTGLRYRRDGDOQLVYVSSITYSKGLGCLEN PGFCGRDVNDHKLDGSPAGDKBEELGNMQTDWDGEQKRVQNVNFPSCDSLOSMSG PGFCGRDVNDHKLDGSPAGDKBEELGNMQTDWDGEQKRVQNVNFPSCDSLOSMSG |

```

GHAYVPAHEPRHCFRLSNACPSPNSKYSILCANPCPDYCHSGFSGMFCSDRCVEACEBNP
GFVSLGELICIPSGOCCGLHPAGSYFKKQCEKVELCVSESNVNNIRCOQPMACRQO
EFGGQGDIIYGGHGAATCTASGDPHATLFTDGLAHHMGTCITVTLRPPCMRSQDSY
FVSVATNNRGGILIEVSYIKAVHTVPELTSLILRGCMVLMGHRVALPVMLAOGRTVY
IRLSNVLITLTPNGLQVRYDGHSLVETLTPSSYGGQVLCGGEFPHLNGKATNST
RYTGVYNNNSLDDNIPDRKTLAGDSMOIGAMWKPPESECEGTFYGGKPSQCSQNSMA
DAMKNKALLINIPQSPFSQCHQVYPPQSSFSACVYHGGCITRGDTLALCRSLQATASLC
AOGAGAPMNRNITFECECPGGSSYSPGSPCPDVTCS11NNRDCPKALPCASETECK
GHILGSLGCVPLQAGOCGCDVPAGSYSPHPEVERNYTENTCRLCTCSVHNNITCFQSTCKP
NOICWALGDLIRASRGVGCVDLPESHYVEFSDSNHSIPACTLVAVKCHPMALP
LEKISAKHEKEGGTEAFRLHEVYIDMVAOVTLQKGRVYINSKVYTLPIISOIPGV
YKSSSITITVNIKRTIGVOVKFQDGNHLEIETLPTTYKQVGLVYINSGVYVYVYVYVY
DEKLANSSSEFVNSWKDDIDIPSCQSLVYDEQOIIPAEQENPSCNCAADIRAREK
EALAPVPAWAOASRIDLPFLVDCANTLCFEGGLYQALCALQAFATCAQSOQLRPP
LMRNSFPGHECALRIDPVPKPPPLCPBESLTLPLVPLDCPAVSYSTNCLPSCSPS
CWDLDGREGCAKVPSCAEGCICQGYVSEFDNSHISIPACLVAVKCHPMALP
SCCTKCYCTGGAIQCGDPRCPGSCGHLCTSDNSNSNCVSKSCQSYVGGPRLFTD
GSEYFLQCRMYVLVIKTLVDVLECEYFPLVGBRKNMDPRRSITLQEVITTVYGVYKQ
IQAGLELVYVRAQGP"
312. 335
misc_feature
/note="89.000, shadow-excellent"
complement(327. 435)
/note="99, 1:xyz, 1. 109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, Length, =,
1090, P. =, 2.3e-37, S. =, 99"
/note="Region: Data base match"
complement(327. 435)
/note="99, 173:xyz, 1. 109, of, GB|U83191|HSU83191, Human
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, Length, =,
1090, P. =, 1.1e-32, S. =, 99"
/note="Region: Data base match"
complement(327. 435)
/note="99, 1:xyz, 1. 109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, Length, =,
1090, P. =, 2.3e-37, S. =, 99"
/note="Region: Data base match"
complement(328. 435)
/note="100, 3614:xyz, 1. 106, of, GB|U83191|HSU83191_1,
Homo, sapiens, Eukaryotae;, mitochondrial, eukaryotes;,
Metazoa;, Human, zonadhesin, (ZAN), mRNA, partial, cds;,
Sperm, membrane, protein., 4/97, Length, =, 363, P. =,
1.9e-14, S. =, 100"
/note="Region: Data base match"
complement(328. 435)
/note="81, 6:xyz, 5899. 6145, of, GB|U40024|SSU40024_1,
Sus, scrofa, Eukaryotae;, mitochondrial, eukaryotes;,
Metazoa;, Sus, scrofa, zonadhesin, (Zan), mRNA, complete,
cds, The, amino, acid, sequence, predicts, seven,
extracellular, domains;, including, a, mucin-like, domain,
and, five, domains, homologous, to, the, D-do. ' ' '
11/95, Length, =, 2476, P. =, 0.0e+00, S. =, 81"
/note="Region: Data base match"
complement(331. 578)
/note="24.00, 82aa,
SYTGVGCPRLNFDGFSYLRGMRVYLKTYDVLDEGVPLLVGBRKNMDPPRSSS
IFLQEVITTVYGVYKQIQAQLELV"
misc_feature
/note="331. 578"
complement(331. 578)
/note="1.000, 0.62, 0.44, 0.40, 122, 0.54, 0.57, 0.60"
complement(339. 588)
/note="7.43, 1ATRA:, 367, CAP:, 337"
complement(345. 561)
/note="98, prediction:, 0%, 98%"
complement(403. 570)
/note="80, 3119:xyz, 1. 166, of, GB|U83190|NMU83190_1,
Mus, musculus, Eukaryotae;, mitochondrial, eukaryotes;,
Metazoa;, Mus, musculus, zonadhesin, (Zan), mRNA,
partial, cds;, Sperm, membrane, protein., 4/97, Length, =,
513, P. =, 2.0e-24, S. =, 80"
/note="Region: Data base match"
complement(475. 567)
/note="70, 3642:xyz, 856. 946, of, GB|U941321|HUM1MUCB_1,
Homo, sapiens, Eukaryotae;, mitochondrial, eukaryotes;,
Metazoa;, Human, mucin, 2, (MUC2), mRNA, sequence, 1/95,

```

| | | |
|----|----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | misc_feature | Length =, 984, P =, 3.2e-14, S =, 70° /note="Region: Data base match" complement(475). .567) /note="70, 1807:YZZ, 13441. .13531, of, GBP L2198 HMMUC2X_1, Homo, sapiens, Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Homo, sapiens, intestinal, mucin, (MUC2), mRNA,, complete, cds... 1/98, length, =, 5179, P, =, 1.5e-30, S, =, 70° /note="Region: Data base match" 606. .903 /note="SINE/Alu, pos., in, rep.: 1. .296" /rpt_family="AlusX" 1004. .1038 /note="83..30, %, 22, length:, 2, #: 18, seq:, tt" /rpt_type=randem /rpt_unit=score: complement(1021. .1313) /note="SINE/Alu, pos., in, rep.: 3. .300" /rpt_family="Aluo" complement(1374. .1684) /note="SINE/Alu, pos., in, rep.: 1. .302" /rpt_family="Aluo" 1745. .2050 /note="SINE/Alu, pos., in, rep.: 1. .302" /rpt_family="Aluo" |
| | repeat_region | |
| | repeat_region | |
| | repeat_region | |
| | promoter | /note="Region: GC Content" complement(2172. .2421) /note="5.62, TATA:, -1, CAP:, -1" complement(2314. .2458) /note="0.850, 0.45, 0.46, 0.56, 1.11, 0.49, 0.54, 0.53" complement(2314. .2458) /note="94.000, excellent" complement(2354. .2464) /note="75, 26:xyz, 5749. .5857, of, GBP U40024 SSU40024_1, Sus, scrofa, Eukaryotae;, mitochondrial, eukaryotes;; Metazoa; Sus, scrofa, zonadhesin, (zan), mRNA,, complete, cds; The, amino, acid, sequence, predicts, seven, extracellular, domains,, including, a, mucin-like, domain, and, five, domains,.homologous, to, the, D-do, .. ,' |
| | misc_feature | |
| | promoter | |
| | exon | |
| | exon | |
| | misc_feature | |
| | Query Match | 28.3%; Score 713.4; DB 11; Length 227968; |
| | Best Local Similarity | 99.9%; Pred. No. 2.1e-107; |
| | Matches 714; Conservative | 0; Mismatches 1; Indels 0; Gaps 0; |
| QY | 1802 gcggagtggaagtctacttccttcccaagaagtgatgcacgacgcgatcccggtcgcga | 1861 |
| Db | 154517 CAGGTTGAGGTCTCACTTCCTTCCACAGATCGTGCCAGCCGCATCCCCCTCCGCCA | 154576 |
| QY | 1862 catctcataggscgtlbgagaccacaagcttggggccctctgcgcgcacacbtgcgctgct | 1921 |
| Db | 154577 CATCTTCATGGGGCCGTGGAGACAACAGCTGGGGCCCTGCGACCACTCGGCTGCT | 154636 |
| QY | 1922 gggtccaacaagctccggagaccccgggggacacccctcccaattgcttccaagagagcgc | 1981 |
| Db | 154637 GCGGCTCCAAACACTCGGGACCCC GGGGGCCACCTCCTCACCTTGCTTCAGGAAGCCG | 154696 |
| QY | 1982 ttctcggcgtcagctagacctgtcacctgtagcgtltgcaaggcgacccaatgccttag | 2041 |
| Db | 154697 TTTCGGGGGTACGATGACCCTGCTCACGTGACGCTGCAAGGGGACGCAATGGCCTTAG | 154756 |
| QY | 2042 cggggatgtctggaacattgataacaactctgaagccctggggatcctacatcccggt | 2101 |
| Db | 154757 CGGGGATGCTTGGAACATTGATMAAACATTTCGAGAGCCCTCGGGGATTCCTCATATCCCT | 154816 |
| QY | 2102 ccccccaatcaagaagctcctctgctctcgcttgaaatgaattaaaggttaagggatgggctc | 2161 |
| Db | 154817 CCCCCAGTCAAAGCTCCTCTGCTCTCGCTTGTAATGATTACGGGTCACGGAGGTGGCTC | 154876 |
| QY | 2162 agagtcacacctcatctgtgtcataattctcattaacccctacacatctctccatcgagac | 2221 |

```
Db 154877 AGAGTCCACCTCTGATTCGATTCATATCCCTACACATCTCTCCACGAGCC 154936
Qy 2222 CCAAGACCCAGCAGAGATATCCACACCCAGCCCTGAGTGTAGCTGACCTAACTG 2281
Db 154937 CCAGACCCACACAGATATCCACACCCAGCCCTGAGTGTAGCTGACCTAACTG 154996
Qy 2282 ACAGTCACTGCTGATTAATCAGAGATAGCATCTTCATATCAGACCCCTTCCTCT 2341
Db 154997 ACAGTCACTGCTGATTAATCAGAGATAGCATCTTCATATCAGACCCCTTCCTCT 155056
Qy 2342 TCTGGGGCTCTCATACCTAGAGACCACTCTGGAGGTTTGTAGCCCTGGAGACTGGC 2401
Db 155057 TCTGGGGCTCTCATACCTAGAGACCACTCTGGAGGTTTGTAGCCCTGGAGACTGGC 155116
Qy 2402 CAGCTCTGTAGTGGAGAGATCGCTGACCATAGCTTATGCAAGAGTGTCTGT 2461
Db 155117 CAGCTCTGTAGTGGAGAGATCGCTGACCATAGCTTATGCAAGAGTGTCTGT 155176
Qy 2462 GGTGAAGAGGCGTGGAGTTCATATCAATAAACCACCTGATATCAATAAGCCA 2516
Db 155177 GGTGAAGAGGCGTGGAGTTCATATCAATAAACCACCTGATATCAATAAGCCA 155231

RESULT 3
HUMYR50H04 580 bp mRNA PRI 29-AUG-1998
LOCUS Homo sapiens full length insert cDNA clone YR50H04.
DEFINITION AF085928
ACCESSION AF085928
NID g3483269
VERSION AF085928.1 GI:3483269
KEYWORDS FLI_CDNA.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 580)
Moessner,J., Tan,F., Maria,M., Kucaba,T., Yandell,M., Martin,J.,
Math,G., Bowles,L., Wylie,T., Bowers,Y., Steptoe,M., Theising,B.,
Geisel,S., Allen,M., Underwood,K., Chappell,J., Person,B.,
Gibbons,M., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,
Schurk,R., Ritzer,E., Kohn,S., Swaller,T., Behrmer,K., Hillier,L.,
Wilson,R. and Waterston,R.
Full Clone Sequencing of the Longest Available Member from Each
Unigene Cluster
Unpublished
2 (bases 1 to 580)
Waterston,R.
Direct Submission
Submitted (24-AUG-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:est@wustl.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No
attempt has been made to verify whether this corresponds to the
full-length of the original mRNA from which it was derived. We
have tried to obtain double-stranded, or double chemistry sequence
across the entire clone, but potentially, there are areas in the
sequence where this level of coverage was not achieved.
Nevertheless, we are confident of the accuracy of this sequence as
all regions of low quality, as defined by PHRAP (P. Green, in
preparation), were visually inspected and edited accordingly. The
consensus quality values for this sequence have been submitted
separately.

SIMILARITY INFORMATION:
probable match to Homo sapiens protein AF053356 (PID:g3135312)
```

```
transferrin-receptor2
The location of this clone is unknown.
FEATURES
Source Location/Qualifiers
1..580
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="Soares_fetal_liver_spleen_1MFLS"
/clone="IMAGE:208759"
1..510
/note="probable match to Homo sapiens protein AF053356
(PID:g3135312)"
misc_feature 43..549
/note="similar to Homo sapiens protein P02786
(PID:g136378)"
misc_feature 43..549
/note="similar to Rattus norvegicus protein A34549
(PID:g112425)"
misc_feature 43..549
/note="similar to Mus musculus protein S29548
(PID:g346962)"
misc_feature 43..549
/note="similar to Cricetus griseus protein A48592
(PID:g539790)"
BASE COUNT 120 a 198 c 168 g 94 t
ORIGIN
Query Match 21.8%; Score 550.4; DB 11; Length 580;
Best Local Similarity 98.9%; Pred. No. 1,7e-80;
Matches 554; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 186 gccgcctgactcaggacatcgcgcgcgcctccgcagaaagctgagaccagctgtg 245
Db 1 GCCCCTGATGATGAGACATTCGCGCGCGCTCCGCCAAGACTGAGACCATGTGG 60
Qy 246 accagacagcactcgtggagctcgaattcccgatccggtcccaaacctgac 305
Db 1 ACCGACAGCACTACGTGGGGCTGCAATCCGGATCCGGCTCACCCCAACCTGAC 120
Qy 306 ttggtcagtagagccgggaagctcgaagagcagctgcctcgtggaagaccctgactac 365
Db 121 TGGTGATGATGAGCGCGGGAAGCTGGAAGACGAGCTGCCGTGGAGAGACCTGACTAC 180
Qy 366 tggccctacagcgcacatcgcaacgtcagggagagctgtgtacgcgccactaagggcgg 425
Db 181 TGCCCTACAGCGGCATCGCAACGTACGGGAGAGCTGTGTACGCCCATACGGCGG 240
Qy 426 ccggaagacctgagagactcggcgagggggtgtgagtcagggggcgctcgtg 485
Db 241 CCGGAAGACCTGAGAGACTCGGGGCCAGGGGCGGTGATCCAGTGGGCGCTGCTG 300
Qy 486 gtgcgctgggggtgcatcagcttcgccaggaagtgacaaatgctcaggaactcgggct 545
Db 301 GTGCCCGTGGGGGTGATAGCTTCGCCCAAGAGGACATGCTCAGAGATTGGGGCT 360
Qy 546 caagagtgctcatataaccagagccagcggactctccagagaccaccagccaagc 605
Db 361 CAAGAGTGTCAATATACCCAGAGCCAGGACTCTCCAGAGACCCCAAGCAAGC 420
Qy 606 ctgtcagccagcagggaggtatagagatatgcaactcgggaactgtgagaccctaacca 665
Db 421 CTGTCCAGCCAGCAGGCGTGTATGGAATGTGACCTGGGGAACCTGGAGACCCCTAC 480
Qy 666 cctggctccctcccttcaatcaaacccaggtccctccagctgacatcaagggctcc 725
Db 481 CTGTGCTCTCTTCTTAATCAACCAACCAAGTTCCCTCCAGTTGATCATCAGGCTTCC 540
Qy 726 agcatccagccagcccat 745
Db 541 AGCATCCAGCCGAGCCCT 560
```

[illegible]

| | | | | |
|----|------|----------------------------------------------------------------|-------|-----|
| Db | 917 | ATGCCAAAGAGGACGAGGACGCTGGAGTCTCATGTACT | ----- | 957 |
| Qy | 586 | aggaaccaccacaagccaaagcctgccagcaagcagtgatgacatgycacctgg | 645 | |
| Db | 957 | -GGATTCACTCATATTATGGAAATACAGATACATTATCCATTCCGACATGCCACCTTG | 1015 | |
| Qy | 646 | gaactggagaccctcacacaccttgccttccttcctcaatcaaaaccagttccctcag | 705 | |
| Db | 1016 | GAACGTGAGACCTTTACACCCCGAGGCTTCCTCTGTTCAACACACCCAGTTTCCACAG | 1075 | |
| Qy | 706 | ttgcatacagagccttcccaagatccagcccacgtcagtgcaagatgtctccc | 765 | |
| Db | 1076 | TTGATCTTCAAGGACTACCCCAATTGCTGTAGACATCTGTAGCACTGTAGCACTGACGACCA | 1135 | |
| Qy | 766 | gaccttgaagaaactcaaaagcccttggcccccaaatgycagggagccctcag | 825 | |
| Db | 1136 | GGCTTTCAGCAAAATGGATGGAGACACATGCTCTTAAGAGTTGGAAGGTGCATTCATT | 1195 | |
| Qy | 826 | gct-----cccttacccttggccccggccagacgtgcgtcagtgatgyltcaacaac | 879 | |
| Db | 1196 | CCTGTAAAGTGACAAACAAGCAGAGGACGATATGTGTGAAGTATGTAACAAT | 1255 | |
| Qy | 880 | acaggaacctccaccatacaaacatcttgcgtgcatactgaagccgcttaagccag | 939 | |
| Db | 1236 | CCATGAAAGACAGGAAAGATTCTGACATCTCCGTCCTTCCAGGATTTGAAAGAACTG | 1315 | |
| Qy | 940 | atcactacgttgatctggggcccaaggagatgcatggygccccagagacataatcg | 999 | |
| Db | 1316 | ATCGGTATGTTGTATTGGAGCCAGAGAGACTCTGGGGCCCAAGAGTGGCTTAAGCTG | 1375 | |
| Qy | 1000 | ctgtggagagcgcataactctctgagcttggtygcgaccttccctcaatgyltgagcaac | 1059 | |
| Db | 1376 | GCACGGAACGTCTATATGTTGGAACTTCCCGTGTGATCTCAGACATAGGAAACG | 1435 | |
| Qy | 1059 | --gcttcggccccgaagaagctctctctcaactgcctggagagtgtygtacttgaa | 1116 | |
| Db | 1436 | AGGGCTTAAACCCAGGCGAAGCATATCTTGTCTGCTGAGGTGCAGGAGACTCAGGAG | 1495 | |
| Qy | 1117 | gcgtggagctcaagagtggtctgaagagctaacagtgctgtcacctcaagagccgtg | 1176 | |
| Db | 1496 | CTGTGGTGTACTGAATGGCTGGAGGGGTACTCTCCATGCTGCATGCCAAGCTTTCA | 1555 | |
| Qy | 1177 | tgtaactgagccttgacaacgcagctctggyggatgtacaaglttcatgccaagacaagc | 1236 | |
| Db | 1556 | CTTACATCAGCTTGATGTCACAGTCCCTGGAGCAAGCCATGTCAAGATTCTGCCAGC | 1615 | |
| Qy | 1237 | cccttcgacaagctcattatgagatgctctgaagcaggtgatatctcccaaccaagtg | 1296 | |
| Db | 1616 | CTTGTCTGTATATCTCTCTGGGAGATATATGAAAGGGGTGAAGATCCAGCAGCAGTCT | 1675 | |
| Qy | 1297 | ggagaaaccttgaagcaggtggtygttcaacaatcccgctcggatgltgtgagatcc | 1356 | |
| Db | 1676 | CAGAGAGCCTTATTAACAGACTTGGCCAGACTGGGTAAAGCAATTTT----- | 1728 | |
| Qy | 1357 | ggcccttaccatlgacaagagtgctatctctcaagccttltggyagtccttcg | 1416 | |
| Db | 1726 | --CCTTGGGCTGTGATATAGCAGCGTCCCTTCTCTGGCGTACTCAGGAATTTCAAGT | 1783 | |
| Qy | 1417 | tcgagttccttatatgagagcagaccaggtcattccattctcgcacacaagaagaca | 1476 | |
| Db | 1784 | TGCTTTTGGTTTCTACAAATAAAGATGAGGAATATGCTCTCCGACACTAAGGGGTGACA | 1843 | |
| Qy | 1477 | cttatgaaacctcataaagtgctgaagccgctgcgcgcgtyggcccaagccgcyg | 1533 | |
| Db | 1844 | CACGTGGAACCTTATGAGAAATTAATAT-----CTGATGCTCTTCTGCTGCTGCTG | 1891 | |
| Qy | 1537 | cccaactgcagaggcaagctccatccgcgtcagctcagcaagatcgccctgcctcagat | 1596 | |
| Db | 1898 | CAGAAGTAGCTGGACAGCAGCTCTCAAGCTGTGACCATATCATATAGACTCTTCTTGACA | 1955 | |
| Qy | 1597 | tcggcgcgtacgggagcgtgctcctcaaggacatcggaaactcaagatctctcggg | 1656 | |

| | | | | | | | | |
|--------------|------------------------------------------------------------------|-------------------------------------------------------------|------|---------------------------|------------|-------------------------------------------------------------|----------------------------------------------------------------|------|
| Db | 1958 | TCGGAGATACAGTGAAGAAATTACTGGCATACAGAGAGAGACTTTTGCTTACATTAAG | 2017 | misc_feature | 2442..2450 | /note="pot. extracellular glycosylation site (aa 727)" | | |
| Qy | 1657 | accctcaagcccgccgctgacccctgcaagtggtgactcgccgaggggaccatcc | 1716 | BASE COUNT | 1399 a | 977 c 1155 g 1479 t | | |
| Db | 2018 | AAGTGGGGAGCTGGGGCTTGACCTTGAGCTGTTTTTTTCCCGGTGACTTCCAGC | 2077 | ORIGIN | | | | |
| Qy | 1717 | gggagcgaggaagctcgagcagagatctcagctcgagagagagagagcagcagctga | 1776 | Query Match | 11.1% | Score 279.2; DB 9; Length 5010; | | |
| Db | 2078 | GAGCTGTACTGCACTGAGAGAAGACATTCGAACAGTACGAGGAGAACGAGTATCC | 2137 | Best Local Similarity | 54.0% | Pred. No. 1e-36; | | |
| Qy | 1777 | caagcagtcacacgctgacataatgcyggtgaggtcttactctctccagcagcagctgt | 1836 | Matches 780; Conservative | 0; | Mismatches 608; Indels 56; Gaps 8; | | |
| Db | 2138 | GCAAGGCGCTGATGACAGATGATGAGTGAAGTATGACTTCTGCCCCCTATCTCT | 2197 | Qy | 481 | tgcctgctgcgcgctggggtatcacagcttcgcccaggaagtgacatcagcagctcg | 540 | |
| Qy | 1837 | cgccagccgagctcccgcttcgccacatctcagcgtgagcgtgagacacagcctggcg | 1896 | Db | 1027 | TGATGTACAGACGCGGAATAATCATTCTTCAGAGAAAGCTTGCAAAAGCTTAA | 1086 | |
| Db | 2198 | CACCAAAAGATGTCCCTTTTCCGACATCTTGTGGCAAGGCGCCACACCTCGGGA | 2257 | Qy | 541 | gggctcaaggagtgctcatatcaccagaagcagcagcctctccaggaaccaccagaac | 600 | |
| Qy | 1897 | ccctgctggaaccactgagctgctgctgctcacaag | 1933 | Db | 1087 | ATGCAATTGGTGTGTGATATACATGACGACAGATAAAT-----TC | 1128 | |
| Db | 2258 | GTCGTGTGAGACATCTGACGCTGTTGAACCAACAG | 2294 | Qy | 601 | caagcctgctcagcagcagcagcagtgatgtgacatgctgcaccttggagacccct | 660 | |
| RESULT | 5 | | | Db | 1129 | CCATTGTTAACGACGAACTTTTCAATTTTGGACATGCTCATCTTGAGGACAGTGACCTT | 1188 | |
| HSTR | HSTR | 5010 bp | mRNA | | | | | |
| LOCUS | Human mRNA for transferrin receptor. | | PRI | 11-APR-1995 | Qy | 661 | acacacctgctccctccctccatcaaacaccagctccctccagctgctcatcagcgc | 720 |
| DEFINITION | Human mRNA for transferrin receptor. | | | | Db | 1189 | ACACACCTGGATTCCTTCTTCAATCACACTCAGTTTCCACCATCTCGGTATCAGAGAT | 1248 |
| ACCESSION | X01060 | | | | Qy | 721 | ttcccaagcattcccaagcccaatcagtgacagacatgctcccgctgctgaggaagc | 780 |
| NID | 937432 | | | | Db | 1249 | TGCCTAATATACCTGTGCTCAGACATCTCCAGAGCTGCTGCAGAAACAGCTGTTGGATA | 1308 |
| VERSION | X01060.1 | GI:37432 | | | Qy | 781 | tcaaaagc-cctgtgagcccccagaagatgagcagcagcagcctccctcagctccctatcac | 839 |
| KEYWORDS | transferrin receptor. | | | | Db | 1309 | TGGAAGAGACTGTCTCTGACTGTAAGAACAGCTCTTACATGTAGATGGTAACCTCAG | 1368 |
| SOURCE | human. | | | | Qy | 840 | ctgggcccgcggcagcagctgctgctgctcacaatcacaagcagcagcctccacccctc | 899 |
| ORGANISM | Homo sapiens | | | | Db | 1369 | AAAGCAAGATG-----TGAACCTCAGCTGACATGCTGCGAAGAGATAAAT | 1421 |
| REFERENCE | 1 (bases 1 to 5010) | | | | Qy | 900 | aacaacatcttcggctgcatcgaaagcgcgcgcagcagcagcagcagcagcagcagc | 959 |
| AUTHORS | Schneider, C., Owen, M.J., Banville, D., and Williams, J.G. | | | | Db | 1422 | CTTAACATCTTTGAGTTATTAAGGCTTTGTAGAACACATCATGTTGTGATGGG | 1481 |
| TITLE | Primary structure of human transferrin receptor deduced from the | | | | Qy | 960 | gcccaagagatgcatgctgagcccccagcagcagcagcagcagcagcagcagcagc | 1019 |
| JOURNAL | Nature 311 (1987), 675-678 (1984) | | | | Db | 1482 | GCCGAGAGAGATGATGGGCGCCGAGCTGCAAAATCCGCTGAGGACACAGCTCTCTA | 1541 |
| MEDLINE | 85012743 | | | | Qy | 1020 | ctggagctgctgagcagccttccctccatg---tgaagcaagcagcctccgcccgcaga | 1076 |
| COMMENT | Data kindly reviewed (19-FEB-1986) by C. Schneider. | | | | Db | 1542 | TTGAAACTTGCCCGCATGATTCTCAGATATGCTCTTAAGATGGGTTTCAACCCACGAGA | 1601 |
| FEATURES | Location/Qualifiers | | | | Qy | 1077 | agtcctctctcatcagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc | 1136 |
| | 1..5010 | | | | Db | 1602 | AGCATATTCCTTGGCATGTGGAGTGGAGTGGAGACTTTGAGTGGTGGTGCACATGAATGG | 1661 |
| CDS | /organism="Homo sapiens" | | | | Qy | 1137 | ctagaagctacccctcagcagctgctgacccctcaaacgcagtgatgctgagcctgagcaac | 1196 |
| | /db_xref="taxon:9606" | | | | Db | 1662 | CTAAGAGGATACCTTGTGCTCCATTAAGGCTTCACTTATTAATCTGATGAATAA | 1721 |
| | 264..2546 | | | | Qy | 1197 | gcagtgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc | 1256 |
| | /note="put. transferrin receptor (aa 1-760)" | | | | Db | 1722 | GCGGTTCTTGATACAGCAATTTCAAGGTTTCTCCAGCCCATGTTGATACGCTTATTT | 1781 |
| | /codon_start=-1 | | | | Qy | 1257 | gaagagctcctgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc | 1316 |
| | /protein_id="CAA25527.1" | | | | Db | 1782 | GAGAAACAAATGCAAAATGTAACATCC---GGTTACTGGCAATTTCTATATCAGGA-1838 | 1838 |
| | /db_xref="PID:937433" | | | | Qy | 1317 | gtgtgtgtaccacatcccaagctgagcagcagcagcagcagcagcagcagcagcagc | 1376 |
| | /db_xref="GI:37433" | | | | Db | 1838 | -----CAGCAACTGGGCGCAGCAAAATTTGAGAAATCCTATTGACCAAT | 1880 |
| | /translation="MMDQARSFNSLGGPGLSTYRSLARQVDSNSHYEMLAQDE | | | | Qy | 1377 | agtcctatcctcagcagccttggagcagcagcagcagcagcagcagcagcagcagc | 1436 |
| | EMADNNTKANTKPKKCSSTCYGTAVYVFLIGMITYLGYCKVEKTECEBIA | | | | Db | 1881 | GCTCTTTCCTTCTTCTTCTGATTTCTGGAATCCACAGATTTTCTTCTTCTTCTGCGAG | 1940 |
| | GTEPVRREEGEDPFAARRLYMDLKRLEKLDSTDTSTIKLNNNSYVPEAGSQ | | | | | | | |
| | KDNLALYENQREFELKSVMDQHVKTIOVDSAGNSIYIDKNGRLYLVENPG | | | | | | | |
| | YVAASKAATVTKLVANFQTKDFEDLYPPVGSIVYAGKITEAEKVAANESLVA | | | | | | | |
| | IGLVLYMDQTKPIVNAELSEFGAHGIDGPTPGSPSNHQPSPRSAGSPLNIPY | | | | | | | |
| | OTLSRAAEKILFGMGEDCDMDKTDSTGMSKSNVLTYSNLYKELTIINIRY | | | | | | | |
| | IKGFVEPDHYVVGADRDANGPRAKSGVSTALLIKLAQFSDMKVLDKGQPSRSITF | | | | | | | |
| | ASWSAGDFSGVATEMLEGLSLDLKAFYINIDKAVLSTNSKVASPLLYTLLEK | | | | | | | |
| | TMQNHKPVTFQGLYDSSNMAKVEKLTJLNAFPPFLAAGIYAVSFCFCEDDYPL | | | | | | | |
| | GTMVTDIKELIERIPELNVARAAEVAAGQFVILKIDHVLNIDYERNSQLSEFND | | | | | | | |
| | LYNQIRADIKEMGLSLQWLYSARGDFEATSRLLTDFGNAKTDBREVMKLRIDYMAVE | | | | | | | |
| | YHFLSPVSPKESPERHVEWSSGSHITLPALENNKLKRNKNGAFNETLFRNQLATATW | | | | | | | |
| | TIGGAANAISGDWMDIDNEF" | | | | | | | |
| | 456..527 | | | | | | | |
| misc_feature | /note="put. transmembrane segment (aa 63-88)" | | | | | | | |
| misc_feature | 1014..1022 | | | | | | | |
| misc_feature | /note="pot. extracellular glycosylation site (aa 251)" | | | | | | | |
| misc_feature | 1212..1220 | | | | | | | |
| misc_feature | /note="pot. extracellular glycosylation site (aa 317)" | | | | | | | |

QY 1437 gacgacagcgctaccatctctgcacacaaagagagacactatagaacctgcaatg 1496
 Db 1941 ---GACACAGATTATCTTATTTGGGTGACACCATGACCACTTAAGGAACGATTGAG 1997
 QY 1497 gtgctcaagggccgctgcgcgcgtgcccagcgctgcccagctgcgaaggcgagctc 1556
 Db 1998 AGGATTCCG---AGTTGAACAAGTGGCAGACGACGTGCAGAGGTCCGTGCTCAGTTTC 2054
 QY 1557 ctcacgcgcctcaagcagcatgcctctgcgcctgcagctgcgcgcgtctacggggagctc 1616
 Db 2055 GTGATTAAACTAAACCATGATGTTGAATTGAACCTGGACTATGAGAGGTAAACAGCCAA 2114
 QY 1617 gtccctcaagcacaactcgggaaacctcaacaggtctctcggggagactcaaggccgagctg 1676
 Db 2115 CTGCTTTCACTTTGTGTGAGGGATCTGGAACCAATAGACAGACATAAGGAATGGCCCTG 2174
 QY 1677 accctgcaatgggtgctacccgcgcggggagactacatccggcgcggaagagctgcgcg 1736
 Db 2175 AGTTTACAGTGGCTGTATTTCTGCTCGTGGAGACTTCTTCCGTCTTCTCCAGACTTAACA 2234
 QY 1737 caggagagctacagcgtcgcgagagagagagagcagcagctgacacgcatgtacacgtgcgc 1796
 Db 2235 ACAGATTTCGGGAATGCTGAGAAAACAGACAGATTGTGATGAAGAACTCAATGATCGT 2294
 QY 1797 ataatgcgggtgagtgcttaactctccttcccagtaagtgctgcgcagcgactcccgcttc 1856
 Db 2295 GTCATAGAGTGAGATGATCATCTTCCTCTCCCTAGTATCTTCAAAAGAGTCTCTCTTC 2354
 QY 1857 cgcacatcttcacgtcgcgctgagagacacagcgtcggcgccctgctgagacacacgtgcgcg 1916
 Db 2355 CGACATGCTCTTCTGGGGCTCCGGCTCTCACACGCTGCCAGCTTACTGGAGAACTTGAAA 2414
 QY 1917 ctgcg 1920
 Db 2415 CTGC 2418

RESULT 6
 HMMTERR 2826 bp mRNA PRI 14-JAN-1995
 LOCUS Human transferrin receptor mRNA, complete cds.
 DEFINITION M11507
 ACCESSION G339515
 NID M11507.1 GI:339515
 VERSION 1
 KEYWORDS transferlin.
 SOURCE Human, cDNA to mRNA, clone pCDTR-1.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2826)
 AUTHORS McCelland, A., Kuhn, J.C. and Ruddie, F.H.
 TITLE The human transferrin receptor gene: genomic organization, and the complete primary structure of the receptor deduced from a cDNA sequence
 JOURNAL Cell 39 (2 Pt 1), 267-274 (1984)
 MEDLINE 85048936
 COMMENT Draft entry and clean copy sequence for [1] kindly provided by A.McCelland, 17-APR-1985. The area coding for the transmembrane domain of the transferrin receptor protein is found at positions 279-362.

FEATURES
 source
 1. Location/Qualifiers
 1..2826
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="3q26.2-qter"
 <1..>2826
 /note="tfrr mRNA"
 gene 96..2378
 /gene="TFRC"
 96..2378
 CDS /gene="TFRC"
 /note="transferrin receptor"

/codon_start=1
 /db_xref="GDB:600-120-433"
 /protein_id="AA61153.1"
 /db_xref="PID:g339516"
 /db_xref="GI:339516"
 /translation="MMDQARSAFSLFGEDELYSTRPSLARQVNDNSHVMKLAUDE
 EEMDNNTKANVTPKRCSSGICCTIAIVIFELIGRMWIGLIGCKVEPTREBRLA
 GTESEVREDEDEPPAARRLWDLKRSLKSDSTPSTIKLKNLSYVREAGSO
 KDELNALYVENQFREFKLKYMRDQHEVKIQKDSQNSVLIYKNGRLVLYENPG
 IYVAASKATVYGLVHANFELSKDFEDLYIPVNSGIYVRAGKITFEAKVANAESLNA
 IGVLIYDQETKEPIVNAELSEFQHAHGTGDPYRGPSEHNHTQPPSRSGLENIY
 OTISRAAEKELFGMEDDPCSDMDTDSICRMVRSKNVLTIVNLVKEIKLIIIFGY
 IKGFVEDPHYVYVNAQBDAMPBGAAGVALLIKLAQMSDVIKADRPSSITF
 ASWSAGDGSVGAETWELGILSLHLAFTTINDKAVLGSNFKVSAFLVLTLEIK
 TMQNVKHPVGTQFLYDQSNMWSKVEKLTLDNAEPFLAVSGIPAVSFCFEDTDPYL
 GTTIDTYKELIERLPELNTKYARAALVAGQVVIKLTIDVELINDYERINSQLSFVRD
 LNOYRADIKEMGLSLQMLYSARQDFPATSRLTDFGAEKTRDFVKMLNDRYMRE
 YHFLSPVSPSESPFRHVFWMGSGSHLTPALLENLKLRKNNNGAFNEFLFRNQLATW
 TIOGAANLSGDVVDIDNEF"
 BASE COUNT 793 a 568 c 676 g 789 t
 ORIGIN 449 bp upstream of PstI site.

Query Match 11.1%; Score 278.6; DB:10; Length 2826;
 Best Local Similarity 55.0%; Pred. No. 1.4e-36;
 Matches 716; Conservative 0; Mismatches 349; Indels 38; Gaps 7;

QY 622 cagtgatagaaatgtgacacttggaacttgagacccctacacactgtctccctct 681
 Db 982 CATTTCTTGGATATCTCATCTGAGGAGACAGTGACCTTACACACTTGATTCCTTCT 1041
 QY 682 tcaatcaaacaggtctccctcagttgcatcatcaagccttccagcatcccaagccagc 741
 Db 1042 TCAATCACACTGACTTCCACCATCTCGTCAATCAGATTGCTTAATATACCTGCCAGA 1101
 QY 742 ccataagtcagacatctccctccgcctgcgtagagaaactcaagg-ccctgtgcccc 800
 Db 1102 CAATCTCCAGACTCTGCAAAAAGCTGTTGGGAATATGGAAGAGACAGTCCCTCTG 1161
 QY 801 caagaatgcaaggagagcctcctcagctccctctacacacttgagcccgccagcagctg 860
 Db 1162 ACTGGAACAAACACTCTCATATGTAGATGATGTAACCTGAGAAAGCAAGATG-----TG 1214
 QY 861 cggctagtgtaacaatcaatcaagagactccaccctcaacaacacatctcgtgcac 920
 Db 1215 AAGCTCAGTGTGAGCAAGATGCTGGAAGAGATMAAAATCTTCAATCTTGGAGTAAAT 1274
 QY 921 gaagcgcctcagagacagatctatcgttcacgcggggcccaaggatgacatggc 980
 Db 1275 AAAAGCTTTGTAGAACCAAGATCACTATGTTGAGTTGGGGCCACAGAGATGCATGGGGC 1334
 QY 981 ccaggagcagctaaatccgctgtgaggagcgtatactccttgagagctgtgagcactt 1040
 Db 1335 CTTGAGAGCTGCAAAATCCGGGTGTAGCAAGCTCTCATTTAAACTTGCCAGATGTGTT 1394
 QY 1041 tcttccatagg---tgagcaagggctccggccgcagaaagtctccttccatcagctgg 1097
 Db 1395 TCAGATATGGCTTTAAAGATGAGTTTCAGCCACAGAAAGCATATCTTTGCCAGTTGG 1454
 QY 1098 gacggtgtgagcttggaagcgtggagcctcagagagtggtcagaagagtcagtcagctg 1157
 Db 1455 AGTGTGAGAGCTTGTGATGGTGTGGTGCACATGATGGCTTAGAGGATACCTTTCGTCC 1514
 QY 1158 ctgacacctaaagccgtatgtagcttgagccttgagcaacgcaagctctgaggatgatacaag 1217
 Db 1515 CTGCAATTAAAGCTTTCACTATATTAATTCGATTAAGCGGTTCTTGATACAGCAAC 1574
 QY 1218 ttcatgccaagaacagcccccttctgaacagctcatctgagagatgctcgaagcagtg 1277
 Db 1575 TTCAAGGTTTTCGCCAGCCACTGTTGTATACGCTTATTAGAAAAACAATGCAAAATGTG 1634
 QY 1278 gatctcccaacacagagtgaggagactctctatgcaaggggtgtgttcaacatccagc 1337

Db 1635 AAGCATCCGGT---TACTGGGCAATTCTATATACAGA-----CAGC 1673
Qy 1338 tggagatgtagtgatgacccgccccctaccatgacagcagtgctattctctcaagcc 1397
Db 1674 AACTGGGCCAGCAAAAGTTGAAACTCATTAAACATGCTGCTTCCCTTCTTGA 1733
Qy 1398 ttgtggagatccctgctgctgagttctctcttaaggagacagcagccctaccatc 1457
Db 1734 TATCTGTGAATCCACGACGATTCTTCTTCTTTTGGCGAG---GACACAGATTATCTTAT 1790
Qy 1458 ctgcacacaaagagagacactatagaaacctgacataagtgctgcgaagccgctgcc 1517
Db 1791 TTGGGTACACCATGACACACTATAGAACTGATGAGAGATTCCG---AGTTGAC 1847
Qy 1518 ggcgtggcccaagccgtggtgcccagctgcgaagcagctctctcatcgcgctcaagcagat 1577
Db 1848 AAATGTGCACAGCAGCTGCAGAGAGCTGCTGCTGATGATTAACCTAACCCATGAT 1907
Qy 1578 cgctgtgccccctcagcttgcgcgttaagcgagcgctgcgtccctcaagcacatcgaggac 1637
Db 1908 GTTAATTGAACCTGACATATGAGAGCTACACAGCCACTGCTTTTCAATTTGTAGGGAT 1967
Qy 1638 ctcaacgagttctctggagacctcaagcccgcggtgcagctgcagctgagtgactcg 1697
Db 1968 CTGAACCAATACAGACAGACATTAAGGAATGGGCTGAGTTTACAGTGGCTGATTTCT 2027
Qy 1698 ggcgtggggagctacatcccggtggtggaagagctgcgcgagagatctacagctcgag 1757
Db 2028 GCTGTGTGAGACTTCTTCCGCTGCTACTTCCAGACTAACACAGATTTCCGGGAAGCTGAG 2087
Qy 1758 gagagagagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1817
Db 2088 AAAACAGACAGATTGTCATTAAGAACTCATATATCTGTCATGAGAGTGAATATAC 2147
Qy 1818 ttctctccagtaagtgctgcgcagcagcagcagcagcagcagcagcagcagcagcagcagc 1877
Db 2148 TTCTCTCTCCCTACGATATCTCCAAAGAGCTCTTTCGACATGCTTCTGAGGCTCC 2207
Qy 1878 ggagacacacagctggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1920
Db 2208 GGCTCTCACAGCTGCAGCTTTACTGTGAGAACTTGAACCTGC 2250

RESULT 7
LOCUS CRUTRANREC 2460 bp mRNA ROD 09-DEC-1993
DEFINITION Cricetulus griseus transferrin receptor mRNA, complete cds.
ACCESSION L19142
ID 9304528
VERSION L19142.1 GI:304528
KEYWORDS transferrin; transferrin receptor.
SOURCE Cricetulus griseus cDNA to mRNA.
ORGANISM Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Mammalia;
Eutharia; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
1 (bases 1 to 2460)
Collins,J.F., Lai,A., Domingo,D., Fitch,M., Hutton,S. and
Trowbridge,I.S.
YRRF is the conserved internalization signal of the transferrin
receptor, and a second YRRF signal at position 31-34 enhances
endocytosis
J Biol. Chem. 268 (29), 21686-21692 (1993)
94012749 Location/Qualifiers
1..2460
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/cell_line="CHO-K1"
/tissue_type="Ovary"
/tissue_lib="Stratagene catalog #937910"
1..2274
/codon_start=1
/function="endocytosis"

/product="transferrin receptor"
/protein_id="AA03576.1"
/db_xref="pid:g304529"
/db_xref="gi:304529"
translation="MMDQARSAISNLFGEPELSTYRFSLAROVGDNSHVEKLAIDE
EENTDNMKASVSKHRLNGRLCGFIIVAVFELIFGFIYGLGCKRPOKCVRAE
TEGNSSEILIOENIIPSSRLVYADIKLSEKIDAEFPDITKOLSTQSRPBGSDKE
NLAYIENQFPRFKLSKVRDEHYADIKYQKSGSAQNAVITINNGSDLYENRGSQVA
YSKATVSGKLHANFGTKRDEEDLYPVNGSLVIVRAKITEKRVANAQSNALIGV
LIYMDQKFPVVEALSLGHAHLGTDPTTQFPSENHQFPPSSGSLPSIPVQI
SRKAELKQNNETNCPSPMNTDSLCKLESQGINVNLVNNVLEKTRILNLEIYKQ
FEEDRIYVGAORDAMGGAASVGTGLLKLQAFSDMVSRGGRFSRSIIFASM
SAGDFGAVATMELEGYLSLKAFETYNLQKVCTGTFRVSASPLVLTLEKMO
DVHRPIDGRPLTRDSNMISKVEDLSIDNNAFFLAISGIPVASFNCENEDRYEDTN
LDTRYELIKVQPLNKNVRAAEVAQOIFIKLTHDELVDVDMYNNKILSFVELNQ
FRADIKAMELSLOWLSARGDEFRAISRITTFHNAEKTRRVRERINRMKVEYEH
LSPVYSPRESFPHIFWGSQSHLTITVLENLKLKRONSSAFNETLFRNLALATWTIQ
GVANALSGDIDINEF"

3'UTR
BASE COUNT 744 a 487 c 539 g 690 t
ORIGIN

Query Match 10.1%; Score 254.4; DB 12; Length 2460;
Best Local Similarity 54.1%; Pred. No. 1.2e-32;
Matches 712; Conservative 0; Mismatches 566; Indels 38; Gaps 8;

Qy 622 caatgtatgacatgtgacactgtagaactggaagccctacacactgctccctcct 681
Db 878 CACTCTTGGACATGCTCATCTAGAACCTGATGATCAACACCTCGGTTGCATCTT 937
Qy 682 tcaatcaaacccagttccctccagttgcatcatcagcagcttccagcatccagccagc 741
Db 938 TCAATCATCTACGTTCTCCCATCTCCATCAGATCAGGATTCAGTATACCTGTTGAGA 997
Qy 742 ccatcagtgagagatgctccctccgctgctgaggaagct-caaagccctggtgccc 800
Db 998 CGATCTCAAGAAAGCTGCGAAGAAAGCTATTTCAAACATGGAACAACTGTCCTCTTA 1057
Qy 801 caagaaatggcagggagccctcctcctcagctccctcctacactggtgcccggcagcagctg 860
Db 1058 GTTGGAACACAGATCTTTGTAGTAGTAGATCTTCA-----CAGGGTATATAAGTGT 1110
Qy 861 cggctagtggtcaacaatcacaagacctcaccacctcaacaacatcctcgtgcatc 920
Db 1111 AACCTCTCTGTGAACAACGCTGTGAAGAAACAAGAACTTAAACATCTTGGGATAT 1170
Qy 921 gaagggccgtccagagccagatcaactggtgcatcggtgcccagagagatgcatggcc 980
Db 1171 AAAGGATTTGAGGAACCAAGACCGGTACATTTGTAGTAGGTCCCAAGAGATGCTGGGT 1230
Qy 981 ccaagagcaactaaatccgctgtggtgagcgctatactcctgagctggtggtgagcctt 1040
Db 1231 CTGTGCTCTGCAAAATCAAGTGTGGAACAGTCTTCTGTGGAACATGTCGCAAGCATTC 1290
Qy 1041 tccctcagtg--gagcaaggtctcggcccgccgaggaagctcctctcatcagctg 1097
Db 1291 TCAGATATGTTGTTCAAGAGGTGGGTTAAACCCAGAGAGCATTTATCTTGCACGCTGG 1350
Qy 1098 gacgtgtgactttgaaagcgtggtggtccagagagtggttagaagctcctcagctg 1157
Db 1351 AGTCCAGAGACTTCCGAGCTGTGTGGTCCACCGAGTGGTAGAGGATACCTTTCATCT 1410
Qy 1158 ctgcacctcaagccgtagtgcgtgtagcctggaacacgagtgctggtggagatgataag 1217
Db 1411 CTGATTTGAAAGCTTCACTATATATCTGATTAATCTGATTAATCTTGTGATACAGAAC 1470
Qy 1218 ttcatgccaagaccagccccctctgacaagctcatgtagtggtcctggaagcagtg 1277
Db 1471 TTCAAGGTTTCTGCCACCCCTTATTTATATACACTTATTTAGAAAGCAATGACGAGACTA 1530
Qy 1278 gatctcccaaccagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1337

| | | | |
|------------|------------------------------------------------------------------|--------------------------------------------------------------------|-------------|
| QY | 1339 | ggatgctcgaggtgatccggccctaccatcgaacgagtgccattctcctcaagccct | 1398 |
| Db | 1167 | ATTGGATTAGCAAAATTGGAGAACTTTCTCCTTGACATAGCTCATCTCCCTTTTCGTGAT | 1226 |
| QY | 1399 | tgtgtggagatccctcccgctcgaagcttcctcttatgtagaacgacaaagcctccattcc | 1458 |
| Db | 1227 | ATTCAGGAAATCCACAGATTTCTTTCTGTTTGTGAGGATG---AGGACTATACCTTATT | 1283 |
| QY | 1459 | tgcacaaaggaaggacactltagaactgcataaagtgctgcgaagccgcctgcgcg | 1518 |
| Db | 1284 | TGGGCACTAAACTAATATACCTATGAGATTTATATCAGAAATTC---CTAGCTCAGCC | 1340 |
| QY | 1519 | ccgtgtgcccaagcgctgtgcccagctcgcgaagcagctctcatccgcgtcgaagccagtc | 1578 |
| Db | 1341 | AAATGCTTCGTACACAGACAGAGAGTGCGCGGCAGTTCTATTTAACTTACCACATGACA | 1400 |
| QY | 1579 | gctctgtccctcgtacttcggccgctaaagggaagctgtctctcaagcaatccgggaacc | 1638 |
| Db | 1401 | TTGATTTGACCTGTGACTATGAGATGTACACAGCAAACTACTGTCTATTATGAAGATC | 1460 |
| QY | 1639 | tcaagagatctctctgtggagccccaagagcccgcggtgcacctgcagtggtgtactcgg | 1698 |
| Db | 1461 | TGACACAGTTCAAACAGATATAAAAAGATATGGGTCTTAAGTCTACATAGCTGTATTCG | 1520 |
| QY | 1699 | cgcgggggagctacatacccggcgcgcgaaaagctgcgcgaggaagatctacaagctcgagg | 1758 |
| Db | 1521 | CTCGTGGAGACTACTTCCTCGCTACTTCTTAGACTTAACACTGATTTTTCATTAATGCTGACA | 1580 |
| QY | 1759 | agagagagcagcagcttgacaagcattacaaogtgcataatgcgggtgtgaagttcact | 1818 |
| Db | 1581 | AAACAAACAGATTCGTCAAGAGAGGAATAAATGATTCGTATATGAAGGGAGATATCACT | 1640 |
| QY | 1819 | tccctccagatcgtctgcgcagccagctcccgctccgcacatctcatgagggcggtg | 1878 |
| Db | 1641 | TCTCTCAACCCCTATGTATCTCCAGAGAGATCTTCCTTTCCGACACATCTCTGGGGCTCTG | 1700 |
| QY | 1879 | gagaccacacgcgttgggcgcctctgtgacaactgcgcgtctgtctgcgtcccaacagctcgg | 1938 |
| Db | 1701 | GCTCTCACACCTCTCAGCTTTTAGTGGAGAACCTGAGACTTGCTGTGAGAAAATAT----- | 1756 |
| QY | 1939 | ggaaccgcggggagcaactctctcaactgcttccagagagaagccgttccgggtcagctcg | 1998 |
| Db | 1756 | -----CACGTCTTCAATGAAACGGCTCTTCAGAAACCAAGTGTGG | 1793 |
| QY | 1999 | gccctctcacctggagcgcgcgaaggagcgcaatgagcgttggcggagtgctcgtgaaca | 2058 |
| Db | 1794 | CCCTGGCTTACGTGAGCTATTTCAGGGAGATGTGCAAAATGCCCTCTCTGTGTACATTTGGAAATA | 1853 |
| QY | 2059 | ttgataacaactctcga | 2075 |
| Db | 1854 | TTGACATATGATTTTAA | 1870 |
| RESULT | 9 | | |
| MTTRMNA | | | |
| LOCUS | MTTRMNA | 2292 bp | RNA |
| DEFINITION | M.musculus mRNA for transferrin receptor. | | |
| ACCESSION | X57349 | | 29-OCT-1992 |
| NID | 954914 | | |
| VERSION | X57349.1 | GI:54914 | |
| KEYWORDS | transferrin. | | |
| SOURCE | house mouse. | | |
| ORGANISM | Mus musculus | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; | | |
| AUTHORS | Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| TITLE | 1 (bases 1 to 2292) | | |
| JOURNAL | Trombridge, I.S. | | |
| REFERENCE | Submitted (23-JAN-1991) I.S. Trombridge, Salk Institute, P O Box | | |
| AUTHORS | 85800, San Diego, CA 92186 5800, U S A | | |
| JOURNAL | 2 (bases 1 to 2292) | | |
| REFERENCE | Trombridge, I.S., Domingo, D.L., Thomas, M.L. and Chain, A. | | |
| AUTHORS | Unpublished | | |

| FEATURES | source | Location/Qualifiers |
|---------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | | 1..2292 |
| | | /organism="Mus musculus" |
| | | /strain="C57BlxDBA/2" |
| | | /db_xref="taxon:10090" |
| | | /tissue_type="hematopoietic" |
| | | /cell_type="702/3" |
| | | 1..2292 |
| CDS | | /codon_start=1 |
| | | /product="transferrin receptor" |
| | | /protein_id="CAA0624.1" |
| | | /db_xref="PI:954915" |
| | | /db_xref="GI:54915" |
| | | /db_xref="SP:EMBL:062351" |
| | | /translation="MMDQASAFSNLFCGEPDLSYTRSLARQVNDGNSHYEMKLADE EENADNNMKASVRRPKRFNRLCPALALVIFELICMSGYLGYCKRVEKEEYKLA ETEDKDSSETMETEDVFTSRLLTMADLTLSKELNIEFADITKOLSONTPYPRG SOKESDLAYIENDFEHEFSKRYMDHYVYIKVSSIGONMVTIYOSNNDLPES EGYVAFSKPTEVSGSKLHANFGTRKDEELSYNGSLVIRAGEITFAKVNAAOSF NALVILYMDKNKPPVVEADLALFGHALLGTGDPYPPGSPSNTOFPSCSLPNI PVOTISRAAEKLEFGMEGSPARWNISCKLELSONAVKLIYVKVLERLINFE GYKIGYEDEPRIVYVGAORBALGAVAAKSSVTGGLLQVFSNISKDGRPSIS IIFNSWTADGFGAATGATELEGYLSLHLAFYIINDKVLVLTGSMKVASPLLYTL MGMIMODVKKHPVDGKSLYRDSNMISYKEKLSFEDNAAVPLFASGIPAVSFCEDDY PYLQTRLDYETALQKVPOLNOMRYLAEVAGOLIILKTHVELINDLYENKSLISF KMDNOKRTIDRMGLSLQWLYSARGYFATSLITDPENAEKTENFVAREINDRIM KVEHLESPYVSPRESPRFRILFPGSSHTLSALVENKLNKOKITAFNEFLFRNOL ATWITQGVANNAISDINWIDNER" |
| BASE COUNT | 685 a | 439 c 515 g 653 t |
| ORIGIN | | |
| Query Match | 8.5%: | Score 213.4; DB: 12; Length 2292; |
| Best Local Similarity | 52.0%: | Pred. No. 5.6e-26; |
| Matches 763; Conservative | 0; Mismatches 636; | Indels 68; Gaps 10; |
| Qy | 616 | agcagcgatgatgacatgacatgacacccgaggaacccgagaccccaacacacgctcc 675 |
| Db | 887 | ACCTTGCACCTCTTGGACATGCTCTATGAGAACCTGGATCCATACACCTGGCTTTC 946 |
| Qy | 676 | ctctctcaatcaaacccagctccctccagctgcatcatcaacgagcctccagacccag 735 |
| Db | 947 | CTTCTTCAATCACTACTGAGTTCCGCGCATCTCAATCATGACGAGGTGGCTATATACCTG 1006 |
| Qy | 736 | cccgagcccatcagctgacagacattgcctcccgccgctgacgaagctcaagg-cccgctg 794 |
| Db | 1007 | TGCAAAACATCTCAAGACACTGCTGCAGAAAAGCTATTGGAAAAATGCAAGAGACTGTC 1066 |
| Qy | 795 | gccccccaagaatgacagaggagagcctccctagagctccctctaacctgagcccgagca 854 |
| Db | 1067 | CTGCTAGATGGAACCTAATATTCTTCTATGTAAAGCTGGAACTTTCA-----CAGATCAA 1119 |
| Qy | 855 | cgactgcgctagctgctgacacatcaacaggaacctccacccccaacaacaacatctcggc 914 |
| Db | 1120 | AATGTGAAGCTCATGTGTAAAAAGCTACTGAAAGAAAGAAAGATTAATTAACATCTTGGGA 1179 |
| Qy | 915 | tgcacgaagcgctcagagccagataactactacgtgtgcatcggggcccgaggagatgca 974 |
| Db | 1180 | GTTATTAAAGGTTATGAGAGAACCAACCCGTTATGTTAGTAGAGAGCCCAAGAGACCT 1239 |
| Qy | 975 | tggagcccgagga---gagagctaaacccgcttgggagagagctataactctctgagctgtg 1031 |
| Db | 1240 | TTGGGTGCTGTTGTTGGCGGAGATGCTCAGTGGGGAAAGGCTCTTCTTTGAAACTTGGCC 1299 |
| Qy | 1032 | cggacctcttccctccatgtt---gagcaagggcttcggcccgcccgagagagctcccttc 1088 |
| Db | 1300 | CAAGATATCTCTGAGATATGATTTCAAAAAGATGGATTTAACCCAGCAGAAAGATATATCTTT 1359 |
| Qy | 1089 | atcagctcggagcagtgatgtaacttggaaagcgctgggctccacaggaatggctcgaagctac 1148 |
| Db | 1360 | GCCACTCTGACTGCAGCGGAGCTTTGGAGCTTTGGTGCACCTGAGTGCTTGGAGGAGATAC 1419 |
| Qy | 1149 | ctcagcctgctcagcactcaaaagccctatctactctgaacccctgagcaaacgagctctgag 1208 |

| | | | |
|----|------|------------------------------------------------------------------|------|
| Db | 1420 | CTTTCATCTTTGGCACTTTAAAGCTTCTACTTAATTAATTAATTTGGATAAAGTTGCTCTTGGT | 1479 |
| Oy | 1209 | gatgacagattctcatgccaagaccagcccccctctgacaagctcatgtagagtctcgt | 1268 |
| Db | 1480 | ACTGATACCTCAAAAGTTCTTGCGAGCCCTTATATATACACTTATGGAAGAAGTATAG | 1539 |
| Oy | 1289 | aagcaggtggaattctcccaaccacaagtgggcagctctctatagaacagtggtgtcaac | 1328 |
| Db | 1540 | CAAGATGTAA-----GCATCCAGATTGATGGAATAAATCTCTATATAGAG-----ACAGC | 1587 |
| Oy | 1339 | aatccacgctbvgagtgctgtgagtgatccgcccctccatgagacagagtgccattcc | 1388 |
| Db | 1588 | AATTGGATTAGCAAAAGTTTGAATA-----CTTCTCTTGACAAATGCTGCATATCTT | 1638 |
| Oy | 1389 | ttcaaggccttctgtggagatccctgcccgtcagttctctcttattgagacagcacaacc | 1448 |
| Db | 1639 | TTCCTTGCAATTTCTGGAATCCAGCAGTTCCTTTGTTTGTGGAGATG---CAGAC | 1695 |
| Oy | 1449 | taccattctcgcacacaaagggagacacttatgagaacctgataaagtgctgcaagc | 1508 |
| Db | 1686 | TATCTTATTTGGGCACTTAATTTGGATTAATGAGCAATTACACAGAAAGTTT---CT | 1752 |
| Oy | 1509 | cgctgcgcgcgcgycgcgcacagcgctgtccgcacgtccgacgagcagctccatccgctc | 1568 |
| Db | 1753 | CAGGTCAACCAAAAGTGTGCTACAGCAGCGGAGAGTGGCTGGTCACCTCTATTATAACTT | 1812 |
| Oy | 1559 | agccaagatcgctgctgccccctcgacttcgscgctacgggagacgtcgctcccaaggac | 1628 |
| Db | 1813 | ACCATGACGTTGTAATTTGAACCTCGGACTATGAGATGTATACAGAAACTACTGTCATTT | 1872 |
| Oy | 1629 | atcggaaccctcaacgaggtctctcgsgaacctaaagcccgcggtcagacctgagag | 1688 |
| Db | 1873 | ATGAAGATCTGAAACCGATTCAAAACAGATATACGAGATATAGGCTTAACTTAAGTGG | 1932 |
| Oy | 1689 | gtgctactcggcgcgggggagactacatccgcgcgcggaagaagctgcgacgagagatcac | 1748 |
| Db | 1933 | CTGTATTTCCGCTCGTGGAGACTACTTCGCGTACTTCTAGACTTAACAACGATTTTCAT | 1992 |
| Oy | 1749 | agctcggagagagagagcgagcagctgcagcagcatgtataacgctgcataatgcgggtg | 1808 |
| Db | 1993 | AATGCTAGAAAACAAACACAGATTTGTCAATGAGGGAATTAAGACGATATATGAAATG | 2052 |
| Oy | 1809 | gagttcactctcccttcccagtaagctgtgcgcagccgagctcccgcttcgcacacttc | 1868 |
| Db | 2053 | GAGTATCACTTCTCGTGGCCCTATGATCTCCAAAGAGATCTCTTTCGACATATCTTC | 2112 |
| Oy | 1869 | atggcgcggtggagacacacagcgtggcgccctgctgagacactgcgagctgctgcctcc | 1928 |
| Db | 2113 | TGGGGCTCTGGCTCTCACACTCTCTCAGCTTTAGTGGAAACTTAAGAGCTTGCTCAAAA | 2172 |
| Oy | 1929 | aacgcgctcggagaccgcccgsgccacctctccactgcttccagagagagccgttccgg | 1988 |
| Db | 2173 | AATATT-----ACGCTTTTAATGAACCCCTCTTCGA | 2205 |
| Oy | 1989 | cgtaagctagccctgctcactcagctgcagcgtgacaaggggacgcacatgcttagcgggat | 2048 |
| Db | 2206 | AACCAATGGGCCCTGGCTACTTGGACTATTCAGGGAGTGCAGAAAGCCCTCTCGTGATAC | 2265 |
| Oy | 2049 | gtctggaacattgataacaactctcga | 2075 |
| Db | 2266 | ATTGGAGATATGACAATGAGTGTTTAA | 2292 |

[illegible]

| | |
|------------|-------------------------------------------|
| ORGANISM | Unknown. |
| REFERENCE | Unclassified. |
| AUTHORS | 1 (bases 1 to 780) |
| TITLE | Israeli,R.S., Heston,W.D.W. and Fair,W.R. |
| JOURNAL | Prostate-specific membrane antigen |
| FEATURES | Patent: US 538866-A 27-23-JUL-1996; |
| SOURCE | Location/Qualifiers 1..780 |
| BASE COUNT | 197 a 182 c 201 g 176 t 24 others |
| ORIGIN | /organism="unknown" |

| Query Match | 5.3% | Score 158.6 | DB: 5 | Length 780 |
|-----------------------|-----------------------------------------------------------------|-------------------|-----------|------------|
| Best Local Similarity | 56.9% | Pred. No. 5.2e-17 | | |
| Matches 390 | Conservative 0 | Mismatches 266 | Indels 29 | Gaps 5 |
| QY 630 | ggaacatgcaacacctggaagacggaagaccccaacaacgtgtccctccctcaatcaa | 689 | | |
| DB 17 | GGACATGGCCCACTTGGAACTGGAGACCCCTTACACCCAGGCTTCCCTGCTTCAACCAC | 76 | | |
| QY 690 | accca--gttccctccagttgcatcatalcaagagccttcccaagcatcccaagcccaaccat | 746 | | |
| DB 77 | ACCCANNNGTTCCACAGAGTGAATCTTCAGAGTACCCCAATGTGTTTCAGACCATC | 136 | | |
| QY 747 | agtgaagacatgtgctcccgctgctgtggaagctcaaag--gccctgtgcccccaag | 804 | | |
| DB 137 | TCTACCACTGACGACGACGAGGCTGTTCGCAAAATGATGGAGACACATGCTTGANAGN | 196 | | |
| QY 805 | aatgacagaggaagcctcctctaggtccctcttaacctt-----gac | 845 | | |
| DB 197 | NGTTGGAAAGGTGGCATTCACANNNTTCCGTGTAAGTGNACNNNAACAAGCAGGAGANNNN | 256 | | |
| QY 846 | cccgagccacgactgcyggtagttgttcaacaatacacagagcctccaccatcaacaac | 905 | | |
| DB 257 | GCCAGANTAAATGCTGAACACTGATGTGAACAATTCATGAAGAACAGAGAGATTCTGATAC | 316 | | |
| QY 906 | attctgcgctcatctgaagaccgctcagaagccagatcaactacgttgtatctcgggccag | 965 | | |
| DB 317 | ATCTTCGGTGTCTATCCAAAGGATTGGAAMAACTGTACTGATGTGTTGATTTGGAGCCGAG | 376 | | |
| QY 966 | aggagatgatgatgcccacagagagcagctaatccctctgtyggagagcgtatctccctgag | 1025 | | |
| DB 377 | AGAGACTCTGTGGGCCCCAGAGTGGCTTAACCTGGCACTGGAACTGGCTATATTTGTTGAA | 436 | | |
| QY 1026 | ctgtytgcgaccttttccctccatgtgtgagcaac--ggcttcggcccgccagacgaagcttc | 1082 | | |
| DB 437 | CTTGCCGTGTGATCTCAGACATAGTGAATAAACAAGGGCTACAAACCGAGCGAAGCATC | 496 | | |
| QY 1083 | ctcttcatcagctgtggaacgtygtgtaacttbgaaacgttgggctccaagagtgagtagaa | 1142 | | |
| DB 497 | ATCTTTGCTACTGTGAGTGAAGAGACTACGAGCTGTGGTGTCTACTAGTAATGGCTGGAG | 556 | | |
| QY 1143 | ggctacccctcagcgtgcctcacctcaaaacgtagtgcgta--ggcttgaacaacgag | 1200 | | |
| DB 557 | GGGTACTCTGCCATGCTCATGCCAACAACCTTCACTTACATCANNGCTTGGATGCTCCAG | 616 | | |
| QY 1201 | tgtctgggagatgacaagtttcatgccaagacagaccgccctctgacaaagtctcatgaga | 1260 | | |
| DB 617 | TCTCTGGAGCAAGCATGTCAAGATTTCTGCACAGCCCTTCTGTATATGCTGCTGGGA | 676 | | |
| QY 1261 | gtgtccctgaagcaggtggaattcc | 1285 | | |
| DB 677 | GTATTATGAGGGGTGAAGAATCC | 701 | | |

| RESULT 11 | | | |
|------------|-----------|-------------|-----|
| GI4981/c | GI4981 | 250 bp | DNA |
| LOCUS | human STS | SHGC-14834. | |
| DEFINITION | GI4981 | | |
| ACCESSION | GI1131744 | | |
| NID | GI4981.1 | GI:1131744 | |
| VERSION | | | |

KEYWORDS STS sequence; primer; sequence tagged site.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 250)
 REFERENCE MYERS,R.M.
 JOURNAL Unpublished (1995)
 COMMENT
 Contact: Richard M. Myers
 Stanford Human Genome Center (SHGC)
 Stanford University School of Medicine
 Department of Genetics, M-344, Stanford, CA 94305, USA
 Tel: 415/7259687
 Fax: 415/7259689
 Email: myers@shgc.stanford.edu
 Primer A: CTACAGAGCTGCGAGCTCC
 Primer B: AGCTGACCTTAATGTGACGG
 STS size: 139
 PCR Profile:
 Initial incubation: 94 degrees C for 90 seconds
 Denaturation: 94 degrees C for 15 seconds
 Annealing: 64 degrees C for 23 seconds
 Polymerization: 72 degrees C for 30 seconds
 PCR Cycles: 30
 Thermal Cycler: Perkin Elmer 9600
 Protocol:
 Template: 25 ng
 Primer: each 1 uM
 dNTPs: each 200 uM
 Tag Polymerase: 0.05 units/uL
 Total Vol: 10 uL
 Buffer:
 MgCl2: 2.5 mM
 KCl: 50 mM
 Tris-HCl: 20 mM
 pH: 8.3
 Prepared with primer pairs derived from T64983--Merck/UniEST.
 FEATURES
 source Location/Qualifiers
 1..250
 /organism="Homo sapiens"
 STS
 primer_bind 99..118
 primer_bind complement(218..237)
 BASE COUNT 60 a 60 c 67 g 59 t 4 others
 ORIGIN
 Query Match 5.8%; Score 146; DB 13; Length 250;
 Best Local Similarity 83.3%; Pred. No. 7e-15;
 Matches 214; Conservative 0; Mismatches 33; Indels 10; Gaps 4;
 Oy 2256 cctcagtgtagctgacctaataatgacgcatctggttaataagaagtagat 2315
 |||||
 Db 247 CCTGAGGTGAGTGCACCTTAATGTGACGG---CACTAGCGGTAATCAGAGAGTAGCAT 191
 |||||
 Oy 2316 cccttcaatcaagcccttcccttctggtgctccatcaactagagacactctgg 2375
 |||||
 Db 190 CCTTCAATCACAGCCCTTCCC--TTCTGGGGCCTCATACCTAGAGACACACTCTGG 133
 |||||
 Oy 2376 aggttgtagccctgtagcctgagccagctctgttagtggagagatcgctggacat 2435
 |||||
 Db 132 AGGTTT---CTAGCCTTGAGACTGCTGCTGTAACT--GGANAGATCGCTGNCACCA 78
 |||||
 Oy 2436 agcctatgagccaaaggtgtgtctgtgtgaagaaggagctgagattcaataataaa 2495
 |||||
 Db 77 TACCTTAATGCGCAACAGGTGCTGCTGTAAGAGGGCGCTGAGTTTCAATATATATAA 18

Oy 2496 ccactgtatatcaataa 2512
 |||||
 Db 17 CCACCTGATATCTATATA 1
 |||||
 RESULT 12
 LOCUS 123811 540 bp DNA PAT 07-OCT-1996
 DEFINITION Sequence 29 from patent US 5538866.
 ACCESSION 123811
 NID 91603681
 VERSION 123811.1 GI:1603681
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 540)
 UNCLASSIFIED.
 AUTHORS Israeli,R.S, Heston,W.D.W. and Fair,W.R.
 TITLE Prostate-specific membrane antigen
 JOURNAL Patent: US 5538866-A 29 23-JUL-1996;
 FEATURES
 source Location/Qualifiers
 1..540
 /organism="unknown"
 BASE COUNT 152 a 106 c 133 g 147 t 2 others
 ORIGIN
 Query Match 5.0%; Score 125.4; DB 5; Length 540;
 Best Local Similarity 60.2%; Pred. No. 1.3e-11;
 Matches 260; Conservative 0; Mismatches 167; Indels 5; Gaps 3;
 Oy 859 tgcgctagtgtagcaaatcacagacccctccacccaatacaaatcttcgctgca 918
 |||||
 Db 75 TGAAGCTCAGCTGACGAATGCTCTGAAGAAGATAAAATCTTAACATCTTGAGATTA 134
 |||||
 Oy 919 tcgaagccgctcagaagcagatcactacgtgttcacatcggtggcccaagagatcagtg 978
 |||||
 Db 135 TTTAAGGCTTTGTAGAACACAGATCATGTGTGAGTGGGCGCCAGAGAGATCATGGG 194
 |||||
 Oy 979 ccccaagagcagctaaatc-ggctgtgggagcgctatactccctggagctgtggagcc 1037
 |||||
 Db 195 GCCCTGAGCTGCAAAATCNCGGTGTAGGCACTCTCCATTTGAAACTGCCACATG 254
 |||||
 Oy 1038 ttctcccatg---tgaagaacggttccggccccaagaagctcccttccatcagc 1094
 |||||
 Db 255 TTTCAGATATGCTTTAAAGATGGGTTTCAGGCCAGCAAGCATATATTGTGCCAGT 314
 |||||
 Oy 1095 tggagcgtgtgacttggagagcggtggctccagcagagtggtcagaagctacc-tcag 1153
 |||||
 Db 315 TGGAGTGTGTGAGACTTGTGATCGTTGGTGCCACTGAATGCTAGAGGGATACCTTTCG 374
 |||||
 Oy 1154 cgtgtgcacctcaaaagccgagtgtagtgcgtgagcctgcaagaacagctggggatga 1213
 |||||
 Db 375 TCNCTGCAATTTAAAGGCTTCACTTATATATCTGATTAACCGGCTCTGTGATCACG 434
 |||||
 Oy 1214 caagttcatgccaagaacagcccccctctcacaagctccatcttagagtgctccgaagca 1273
 |||||
 Db 435 CAATTTCAAGTTTCTGCGACAGCCACAGTGTGTATACGTTATTGAGAAAACAATGCAAAA 494
 |||||
 Oy 1274 ggttgatctcc 1285
 |||||
 Db 495 TGTGAAGCATCC 506
 |||||
 RESULT 13
 AF009921 AF009921 2570 bp mRNA ROD 30-NOV-1997
 LOCUS Rattus norvegicus ileal peptidase 1100 mRNA, complete cds.
 DEFINITION AF009921
 ACCESSION AF009921
 NID 92653778
 VERSION AF009921.1 GI:2653778
 KEYWORDS

SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 2570)
AUTHORS Shneider,B.L., Thevananther,S., Moyer,M.S., Walters,H.C., Rinaldo,P., Devarajan,P., Sun,A.Q., Dawson,P.A. and Ananthanarayanan,M.
TITLE Cloning and characterization of a novel peptidase from rat and human ileum
JOURNAL J. Biol. Chem. (1997) In press
REFERENCE 2 (bases 1 to 2570)
AUTHORS Shneider,B.L., Thevananther,S., Moyer,M.S., Walters,H.C., Rinaldo,P., Devarajan,P., Sun,A.Q., Dawson,P.A. and Ananthanarayanan,M.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1997) Pediatrics, Yale University, 333 Cedar Street, New Haven, CT 06520, USA
FEATURES
source Location/Qualifiers
1. 2570
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="ileum"
28. 2265
/note="ileal peptidase"
/codon_start=1
/product="1100"
/protein_id="AA87644.1"
/db_xref="PID:2653779"
/db_xref="GI:2653779"
/translation="MHMAKILGVIGAAALGLGIIIGFAIPKATPEPLASSVSDSDLDIALDSYMGOLDASRIRENIRELSKEPHVATSARDELIVLLIGRKMSGSDTA KTYETVLVSFSTEPNSVEVNGVIFVHSPENKLTGEQAPNLYOYAVAP PGTGPIVLVANSSEDDFKLEAGINIKGIALTRGSGVGAALNAARHGVAV LVYTPGDIINDKSLPNEFFPNSGCLPESGVSRYEYFGDLPYLPAPHVSFLD PHNISGFPDIPTOPIGFEDAKMLCNLMSAPDSMOGALGCEYKLPGEFENGNEPA GSEKVSYNRLERNSNVLGIIOGAVPDRVYVIGNRDSVHGVADPSSCTAVIL EISRYLGLTKRKGTWRPRRSITFASMGAREBELISTETPEFLSKIDERTYITIND ISVFSNATIRAGGTTPVOSVITSATREISAPSSGSLSTDNIRITNRSFVYGLVPS MGTLAGSDYASFIHFLGTTSMDLAVTYDRSKTSARIPTTYHTADFEDYVKEFLDPG FSSHQAVARTGASVLRSLDSFLPLNVDSEYTSQSFQAQENIGALBESHNSLIG PLVTAVEKFAAALNQHILTLKSSSPDLQVRVNDQMLLEAFVLPRAFPERY YSHVLMAPRTASVATFPGLANMYARAQELINSAAEMAEVERQLSIAYVALECAATLQ PVTDL"

BASE COUNT 575 a 747 c 690 g 558 t
ORIGIN

Query Match 4.1%; Score 103.6; DB 12; Length 2570;
Best Local Similarity 50.8%; Pred. No. 3,4e-08;
Matches 402; Conservative 0; Mismatches 354; Indels 36; Gaps 5;

QY 902 caacatcttcgcgtgcgaagccgcgtcgaagccagatcaactacgtgtcgcggcgc 961
DB 1083 CAACGTCCTGGGGCATATTCAGGGGCAAGTGCATGATGATGTATGATACGGAA 1142
QY 962 ccaagaggatgcatgggcccccaagagcagctaacccgcgtgtgggaacggtactcct 1021
DB 1143 CCACCCGGACAGCTGGGTACATGGGCGACAGTGCACCCAGCAGAGGCACTGCTCTT 1202
QY 1022 ggggcgtggcggaactttccctcatgtggaacaggtc---cgcccccgcgaag 1078
DB 1203 GGGGATCTCTCGAGCTCTGGGACCTTGCTAAAGAGGACCTGGGCGCCCGGTAGATC 1262
QY 1079 tctctcttcacatgcgtggagcgggtgtgactttggaagcgtgtgcctcaacggagtcgc 1138
DB 1263 CATCATATTTCAGCTGGGGAACAGAGAGGTTCCGGGCTCATGCGCTCCACAAATTCA 1322
QY 1139 agaagcctacatgcgtgtgcacactcaaacgcgtagtgtacgtgtagccttggaacgc 1198
DB 1323 AGAGGAGTCTCTCAGCAGCTGAGAGGCGCACAGTGCATCATCAACGTGGACATCTC 1382

QY 1199 agtgcctggggatgacaaagtttcatgcccagaacacgccccctcttgacaagtcctatga 1258
DB 1383 TGTGTTCTCCAAATGCTACCTCCGGGCACAGGGACACCCCTGTACAAAGGTGATCTT 1442
QY 1259 gagtctcctgaagcagctgtgattctcccaacacagctggcgagactcctatgaacagtt 1318
DB 1443 CCTGCGACCAAGAGATAGTACACACAGCTCCAGTGGCCCTCAGTATATGACAACTG 1502
QY 1319 ggtgttcaacaatcccaagctgtgatgtcgaagttgatcgcgcctcaaccatgagaa--- 1375
DB 1503 GATCCGATATACATACCCGAGTAGTCCAGTATATGAGACTGTTCCAGCATGGGCACT 1562
QY 1375 -----gcagtcctat---tcttcaagcctcttggggagtcctccgctgagtt 1423
DB 1563 GGGTCAGGCGACGACTATGCTTCACTTCACTTCTGGCATTACCTCATGAGACT 1622
QY 1424 ctctcttatgagagc-----gaccagcctaccatcctcctgacaca 1468
DB 1623 TGCCATACACTATAGACCGGAGACAGACACTCAGCCAGATCTACCAACTACACAGAC 1682
QY 1469 ggaagacacttatgagaacctgcataaggtcgtgcaaggcgcctgcgcgcgtggccca 1528
DB 1683 GTTGACACGTTTGACTATGTGAGAGATTTCTGGACCTGGTTTCAGCAGC---CACCA 1739
QY 1529 ggcgcgtggcccaagcctgcgcagcgcagctcctcctcgcgcctcaacagatcgcctgctgc 1588
DB 1740 GGGTGTGGCCCGACTGCTGGGAGTGTGCTGCTCCGGCTGAGTGCACGCTTTCTCTGCG 1799
QY 1589 cctgcagcttcgcgcgtcgaagggagcgtcgtcctcgcagcagcagcagcagcagcagcag 1648
DB 1800 CCTCATATGCTAGTACTATATGAGACACTACAGAGTTTCTCTACAGGCTGCTCAAGAGA 1859
QY 1649 cctctggagacct 1660
DB 1860 CCTTGCGCGCT 1871

RESULT 14
LOCUS HSA012371 2320 bp mRNA PRI 25-MAR-1999
DEFINITION Homo sapiens mRNA for NAALDase L protein.
ACCESSION AJ012371
NID 94539526
VERSION AJ012371.1 GI:4539526
KEYWORDS naaladase L gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2320)
AUTHORS Pangalos,M.N., Neef,J.M., Somers,M., Verhasselt,P., Bekkers,M., van der Helm,L., Fridpols,E., Ashton,D., and Gordon,R.D.
TITLE Isolation and expression of novel human glutamate carboxypeptidases with N-acetylated alpha-linked acidic dipeptidase and dipeptidyl peptidase IV activity
JOURNAL J Biol Chem. 274 (13), 8470-8483 (1999)
MEDLINE 99185063
REFERENCE 2 (bases 1 to 2320)
AUTHORS Pangalos,M.N.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) Pangalos M.N., Biotechnology, Janssen Pharmaceutica, 30 Turnhoutseweg, 2340 Beerse, BELGIUM
FEATURES
source Location/Qualifiers
1. 2320
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q12"
/tissue_type="small intestine"
17. 2239
/gene="naaladase L"
17. 2239
/gene="naaladase L"


```

/function="peptidase"
/codon_start=1
/evidence=experimental
/product="NAALadase L protein"
/protein_id="CAB39968.1"
/db_xref="PID:e1420518"
/db_xref="PID:94539527"
/db_xref="GI:4539527"
/translation="MWTKYGLGLGAAALIGLIGLIGLPAIRKANSAPLDLEI
LETVGQDLNHRILREMLREPHLASSRDELDVLQNRKMDPEPSIDSAEATY
EVLSPSEDOEPNVVDIVPTGGIISCHRETBENIVGEGPDVQVPTAYAPSGTPO
GLIVANRGADEDFEKELOTGKLEGIATLRGGRGAKANAAKHGVALVYTD
PADINDLSSPDETPPNWYLPSPGVRSYVFPDLPVLPVAPSPRVLAVS
GEPDIPOTOPGODARDILCNLTGTLAPATWQALCHYRLGFRPDSGFPADSOVN
VSYNRLELBNSSNVLCITRGAVEPDRTVLYGNHRDSWAGAVDPSSGTAVLLESRV
LGLTLKGTWRPRRSIVFASWGAEEFGLISTEFTFEFKLQERTVAYINVDISVA
NATLRVQGTFFPVDSVVFSAKREIRSPQDLSYDWMIRFNSSPYGLVPSLSIG
AGSDYAFVHFLGISMDIAYIDRSKTSARIPTYTAEDTIDYDKHDPFSSHO
AVARTGSVILRLSDSFELPLKYSSETLRSFLQAOODLGLAIDRHSISLPIVTA
VEKEEAAALGORISDTLOKSPDPLQVRNLNDQMLLETFINPRAFPERYSHVL
MAPSHGIRSHITPELSNACSRARDTASGEAMAEVQRLSTIVTALBGAATLNPVADL
"
BASE COUNT      453 a      725 c      671 g      470 t      1 others
ORIGIN
Query Match      3.68; Score 91.6; DB 10; Length 2320;
Best Local Similarity 46.68; Pred. No. 3.1e-06;
Matches 672; Conservative 0; Mismatches 674; Indels 96; Gaps 8;
QY 314 tgaagccggaaggtcgagagagcagctgcgctgagagagaccctgactctatcccta 373
    |||||
Db 415 TGAGGAGACCTGACCCGGGGAGCAAGGGGGCCAGATGCTTACACCTATGCTGCTTA 474
QY 374 cagcgccatcggaacgctcaacgagagctgtgtacgcccactacgagcgccgaaga 433
    |||||
Db 475 TGCTCTTTGGAGACCCGAGGGCCCTGCTGTATGCCAACCAGGGGGGGAGAAAGA 534
QY 434 cctgcagagacctgcgagcgcaagggcgtggatccagctgagcgccctgctgtgctg 493
    |||||
Db 535 CTTTAAAGAGCTACAGACTCAGGGCATCAAACTTGAAGGCCACCAATTGCCCTGACTCGATA 594
QY 494 gggggtgatcagcttcgcccgaaggtgacaaatgctcagaacttcgggggtcgaagagt 553
    |||||
Db 595 TGGGGGCTGAGGGCGGTGGGCCAGAGGCTGTAGACGCTGCCAAGCAGGGGAGCTGGGCT 654
QY 554 gctcataaccagagcagcagcagcttcaccagagaccaccaagcgaacgctgtccag 613
    |||||
Db 655 GGTGTGACACAGACCCCTGCCGACATCAAGATGGGCTGAGCTACACCCGAGAAACCTT 714
QY 614 ccagcagcagctga-----tggaatgtgcaactggaactggagac 656
    |||||
Db 715 TCCCACTCTCGTACTGCTGCCCTCCAGAGTGGAGGAGGAGGCTCTACTACGATATT 774
QY 657 cccatacacactgctcccttccttcataaaccagctccctccagctgac----- 713
    |||||
Db 775 TGGGGACCCCTTGACTCCTTACCTTCCAGCCGCTCCCTTCTTCCGCGTGGACCTTGC 834
QY 713 -----atcagagccttcaccagaccagccaccatcagctgcaagcatctgcctccg 766
    |||||
Db 835 CAATGTCTCCGGATTTCGCCCAATTCTTACACAGCCCTTTGGCTTCCAGATGCAAGA 894
QY 767 cctgcctaggaagctcaaaagccctgtgccccccaagaatgtgcaaggagcctctcag 826
    |||||
Db 895 CCTGCTCTGTAACTCAACGGAAGACTTGGCCCGACCTGCGAGG---GAGCACTGGG 951
QY 827 ctccacctatactgggccccgggccaagac----- 859
    |||||
Db 952 CTGCCACTACAGGTGGGTCCGGCTTCCGGCTGACGAGACTTCCAGAGACAGCCA 1011
QY 859 --tgcgagctagtggtcaacaatcacagagaccctccaccatcaacaacatctcggtg 916
    |||||
Db 1012 GGTGATGTGAGCGCTTACACCGCGCTGGAGCTGAGAACTCTTCCAGAGCTCTGGGCAT 1071

```

```

QY 917 catcgaagccgcgtccaaagccagatactactgctgtatcgtgggcccagagagatgcatg 975
    |||||
Db 1072 CATTCGCGGGGCTGTGGAGCCTGATCCCTGACTGTCTATGGGAACCCAGGACACCTG 1131
QY 977 ggcgccagagagcagcctaaccgcctgtgaggagagctatctccttgagctgtgaggagc 1036
    |||||
Db 1132 GGTCCACGGGGCTGTGGACCCAGACACTGGCACCGCGCTCTCTGTGAGCTCTCCCTGT 1191
QY 1037 ctttccttcacttggtgagcaacgg---cttcggcccccgagaagctcctcttcacag 1093
    |||||
Db 1192 CCTGGGAGACCTGCTGTAAGAAGGACACTGGCGTCTCCGAGATCAATCGTGTGGAG 1251
QY 1094 ctggagcagctgtgtgactcttggaagcgtgggtcccaagagatgtgctagaagctacc 1153
    |||||
Db 1252 CTGGGGGGCTGAGAGTGTGGCTCATTTGGCTCCACCGAAATTACAAAGATGTTCTTCAA 1311
QY 1154 cgtgctcacctcaaaagccgtagtgatcgtgagccttgacaacgcaagtgctgaggagatg 1213
    |||||
Db 1312 CAAGCTCAGAGAGCGCACGGTGGCTTACATCAACGTTGACATCTCGGTGTTGCCAACGC 1371
QY 1214 caagtctcaatgccaagaccagcccccttcgacaagctcattgagagtgctcctgaagca 1273
    |||||
Db 1372 TACCCTTAGGGGTGAGGGAGCGCCCTGTGCAAGCGGTGCTTCTTGCACCAAGAGA 1431
QY 1274 ggtgagatctcccaaccacagctgggcagactctctatgaacaggtgtgttaccacatcc 1333
    |||||
Db 1432 GATCCGCTGACCAAGCCCTGCGACCTGAGCATTCTACGACAACTGATCCGGTACTTCAA 1491
QY 1334 cagctgtgagtgtgagtgatctcgccccctaccatgga-----cagcag 1378
    |||||
Db 1492 CCGCAGACGCCGGGTGACGGCTGTGTCGCCAGCTTGAGTTCTGTGGTGTGCGACAGA 1551
QY 1379 tgcctatcttcacagccttctgtgagagtccttcgctcgaagcttccttataagagga 1438
    |||||
Db 1552 CTATGACACCTGTGTTACCTTCTCGGGCATCTCTCATGAGACATTCCTATACCTATGA 1611
QY 1439 c-----gaccagcctaccatcttcgtcacacaagaagagacactatga 1483
    |||||
Db 1612 CCGGAGCAAGACTTACACCGAGATCTACCCACCTTACACACACACCTTTGACACCTTTGA 1671
QY 1484 gaactctcataaggtgtctgcaagcgccctgcgcgcgctgtgccaagcggtggccagact 1543
    |||||
Db 1672 CTATGTGACAAAGTTTGTGACCCGGGCTT---CAGCAGCCATAGGCTGGGCCCGGAC 1728
QY 1544 cgcagggcagccttcctcatcgcgtccagccagatcagcgtcgtgccttcgacttcggcgcg 1603
    |||||
Db 1729 AGCGGGAGTGTGATTTCTCGGGCTCAGTGAAGCTTCTTCTGCCCTCAAGTCACTGA 1788
QY 1604 ctacgaggagctgtctcctcagggcacatcgggaacctcaacgagatctcttggaactcaa 1663
    |||||
Db 1789 CTACAGTGAACACTCCGACGCTTCTGTGAGGACGCCAGCAGCAAGATCTTGGGCCCTGCT 1848
QY 1664 gg 1665
    ||
Db 1849 GG 1850

```

RESULT 15
 LOCUS 123810 660 bp DNA PAT 07-OCT-1996
 DEFINITION Sequence 28 from patent US 5538866.
 ACCESSION 123810
 NID 91603680
 VERSION 123810.1 GI:1603680
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 660)
 AUTHORS Israeli, R.S., Heston, W.D.W. and Fair, W.R.
 TITLE Prostate-specific membrane antigen
 JOURNAL Patent: US 5538866-A 28 23-JUL-1996.

| | | | |
|----|------|----------------------------------------------------------------|------|
| QY | 1026 | ctgtgtgagagacctttctctccatggtgagaaac---gcttcgcgcgcgcgaagcttc | 1082 |
| Db | 436 | CTTCCCGGTGATCTAGACATAGTGAATAAACGAGGGCTACCAACCGAGGCGAAGATC | 495 |
| QY | 1083 | ctcttcacagcttgagacggttggtgacttggaaacgcttggtcccaagatggtcctaa | 1142 |
| Db | 496 | ATCTTTCTGCTCGAAGTGCAGGAGGACTACGAGAGCTGTGGGTGCTACTGAATGCGCTGG | 555 |
| QY | 1143 | ggctactcaagcgtgctgcgaacctccaagccgtgttaagtga--gctctggacaacgag | 1200 |
| Db | 556 | GGGATACCTGCCATGCGTCGCAATGCCAAAGCTTCTACTTCATCATANNCTGGATGCTCCAG | 615 |
| QY | 1201 | tgcctggagagatcaaaattccatgccaagaacacagcccccttcctcaaaatccatttga | 1260 |
| Db | 616 | TCTCTGGGAGCAAGCCATGTCAGATTTCTGTCGACGCCCTTGCTGATATATGCTGCGGGGA | 675 |
| QY | 1261 | gtgtcctcgaagcaggtgtgattctcc | 1285 |
| Db | 676 | GTATTATGAAGGGGTGAATAATCC | 700 |

RESULT 2
ID 065537 standard; CDNA; 540 BP.
AC 065537;
DT 11-JAN-1995 (first entry)
DE Sequence used in identification of PSM Ag CDNA.
KW Prostate-specific membrane antigen; PSM; prostate cancer;
KW transmembrane glycoprotein; Imaging; targeting; tumour detection;
KW antibody detection; ds.
OS Homo sapiens.
PN M09409820-A.
PD 11-MAY-1994.
PF 05-NOV-1993; U10624.
PR 05-NOV-1992; US-973337.
PA (SLOK) SLOAN KETTERING INST CANCER.
PI Fair MR, Heston MDW, Israeli RS;
WP1: 94-167129/20.
PT Prostate-specific membrane antigen and DNA encoding it - is
PT useful for detecting haematogenous micro-metastatic tumour cells
PT and for identifying ligands which bind to PSM Ag
PS Disclosure; Page 121; 16pp; English
CC This sequence is given in the specification but is not mentioned in
CC the disclosure. The PSM coding sequence is useful for suppressing or
CC modulating the metastatic ability of prostate tumour cells to grow,
CC or for eliminating them. The protein is useful to identify or purify
CC ligands of the Ag. It is also an attractive target for Ab-directed
CC imaging and targeting of prostatic tumour deposits.
SQ Sequence 540 BP; 152 A; 106 C; 133 G; 147 T;

| | | | | |
|-----------------------|--------------|-------------------|----------------|------------|
| Query Match | 5.0% | Score 125.4 | DB 1 | Length 540 |
| Best Local Similarity | 60.2% | Pred. No. 2.4e-17 | | |
| Matches 260 | Conservative | 0 | Mismatches 167 | Indels 5 |
| | | | | Gaps 3 |

| | | | |
|----|------|--------------------------------------------------------------------|------|
| OY | 859 | tcgcgcctcagtcagttcaacaatccacaggaaccccccacccaatcaacaacatcttcgcgtcga | 918 |
| Db | 75 | TGAAGCTCACGTCAGGCAATGTCTGTGAAGAGATATAAAATTTCTTAACACTTTTGGAGTTA | 134 |
| OY | 919 | tcgaagcgcgcctcagcagccagatcaactaagtctgtcattcgtgggccagagagatcagtcagg | 978 |
| Db | 135 | TTAAAGCTTTGTATGAACCAAGATCACTATCTTTTAATTTGGGGCCAGAAAGATGATATGGG | 194 |
| OY | 979 | ccccagagcagcgttaaatc-cgcctctgtggagcgcatactccctgagcgttgtgtggcacc | 1037 |
| Db | 195 | GCCCTGGAGCTGCATAAATNCGGGTATGGGCACAGCTCTCTCTATTGAAACTTCCCAAGATG | 254 |
| OY | 1038 | ttttcccccctcagtg--tgagcaacagcgtcttcggcccccagagaagttcctctctcatcagc | 1094 |
| Db | 255 | TTTCAAGATATGCTTTAAAAGATGGGTTTCAACCCAGCAAGACATTATCTTTGCCAGT | 314 |
| OY | 1095 | ttggagcgttgctgacctttggaagcgttgagctccacgagatgtgctatagaagctacc-tcag | 1153 |

| | | | |
|----|------|--------------------------------------------------------------------|------|
| Db | 315 | TGAGAGTGTGAGACACTTTGGATGCGTTGGTGCCACTGAAATGGCTAGAGGGATACCTTTTCG | 374 |
| Oy | 1154 | cgtagtcacacctcaaaagccgtagtgcagtgagccctgcgaacaagcagtgctctgaggagatga | 1213 |
| | | | |
| Db | 375 | TCNCTGCAATTTAAAGCGTTTCACTATATATATCTGGATTAAGAGCGTTCTTGATGATCAG | 434 |
| Oy | 1214 | caagtttcacatgcgaagaccagccccctctgcacaagttcattgagatgctctgaagca | 1277 |
| | | | |
| Db | 435 | CACCTTCAAGGTTCTCTCCAGCCCACTGTGTATACGCTTATGAGAAACATCATCAAA | 494 |
| Oy | 1274 | ggtgatattctcc | 1285 |
| | | | |
| Db | 495 | TGTGAAGCATTC | 506 |

RESULT 3
ID T19421 standard; cDNA to mRNA; 93 BP.
T19421
AC T19421;
DT 04-JUL-1996 (first entry)
DE Human gene signature HUMGS00446.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS//) MATSUBARA K.
PA (OKUB//) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 380; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 93BP; 28 A; 17 C; 24 G; 23 T;

```

Query Match          3.5%  Score 89.4;  DB 1;  Length 93;
Best Local Similarity 97.8%;  Pred. No. 4.3e-10;
Matches 90;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

QY  2421  gatcgctgcccataagcctataggcaacagctgctctgtgtgaagggcgctgaat 2480
      |||||
Db    1  gatcgttgccacccatagccttatggcccaacagtgctgtgtgtgaagggcgctgaat 60
      |||||

QY  2481  ttcaatatcaataaaccaacctgatatcaata 2512
      |||||
Db    61  ttcaatatcaataaacccacctgatatcaata 92
      |||||

RESULT  4
Q65536
ID      Q65536 standard; cDNA; 660 BP.
NC      Q65536;

```

| | | | | |
|---------------------------|-------|--------------------|-----------|-------------|
| Query Match | 3.5% | Score 87.2; | DB 1; | Length 660; |
| Best Local Similarity | 55.8% | Pred. No. 1.6e-09; | | |
| Matches 240; Conservative | 0; | Mismatches 181; | Indels 9; | Gaps 4 |

| | |
|--------|-------------------------------------------------------------------|
| RESULT | 5 |
| 065520 | |
| ID | 065520 standard; cDNA; 2653 BP. |
| AC | 065520. |
| DT | 11-JAN-1995 (first entry) |
| DE | Prostate-specific membrane antigen cDNA. |
| KM | Prostate-specific membrane antigen; PSM; prostate cancer; |
| KM | transmembrane glycoprotein; imaging; targeting; tumour detection; |
| KM | antibody detection; ds. |
| OS | Homo sapiens. |

| | | | | | | | |
|-----------------------|-------|--------------|-------|------------|-----|--------|------|
| Query Match | 2.1% | Score | 53.8 | DB | 1 | Length | 2653 |
| Best Local Similarity | 48.8% | Pred No. | 0.013 | | | | |
| Matches | 145 | Conservative | 0 | Mismatches | 152 | Indels | 0 |
| | | | | | | Gaps | 0 |

| RESULT | 6 | |
|--------|------------------------------------------|-------------------------------------------------------------|
| AC | 136785 | |
| ID | T36785 | standard; cDNA; 2654 BP. |
| AD | 04-NOV-1996 | (first entry) |
| DT | prostate-specific membrane antigen cDNA. | |
| DE | prostate-specific membrane antigen; PSM; | promoter; prostate cancer |
| KW | metastasis; gene therapy; diagnosis; ss. | |
| OS | Homo sapiens. | |
| FH | Key | Location/Qualifiers |
| FT | 5'utr | 1..261 |
| FT | | /*tag- a |
| FT | cds | 262..2233 |
| FT | | /*tag- b |
| FT | misc_feature | /product= PSM antigen |
| FT | | 114..380 |
| FT | | /*tag- c |
| FT | | /note="bases 114-380 (-147 to +109) are absent in PSM cDNA" |
| FT | polya_signal | 2352..2357 |
| FT | | /*tag- d |
| PN | W09626272-A1. | |
| PD | 29-AUG-1996. | |

| Query Match | 2.1% | Score 53.8 | DB 1 | Length 2654 |
|---------------------------------------------------------------------------|----------------|-----------------|----------|-------------|
| Best Local Similarity | 48.8% | Pred. No. 0.013 | | |
| Matches 145 | Conservative 0 | Mismatches 152 | Indels 0 | Gaps 0 |
| 23-FEB-1996: U02424. | | | | |
| 24-FEB-1995: US-394152. | | | | |
| 02-JUN-1995: US-470735. | | | | |
| 02-JUN-1995: US-466381. | | | | |
| (SLOK) SLOAN KETTERING INST CANCER RES. | | | | |
| Fair WR, Heston MDW, Israeli RS; | | | | |
| WPI: 96-402365/40. | | | | |
| P-PSDB: W02234. | | | | |
| DNA encoding alternatively spliced prostate-specific membrane | | | | |
| antigen - useful to develop prods. for detecting haemotogenous | | | | |
| micrometastic tumour cells, or prostate cancer progression | | | | |
| Example 1: Fig 4A-D: 284bp; English. | | | | |
| A cDNA clone (T36783) codes for human 100 kDa prostate-specific | | | | |
| membrane (PSM) antigen (W02234), an integral membrane glycoprotein | | | | |
| that is very highly expressed in prostatic tumours and metastases. | | | | |
| It was obtd. from lymph node carcinoma of prostate (LNCaP) cell | | | | |
| mRNA by PCR amplification (see also T36795-808) and screening of an | | | | |
| LNCaP cDNA library using an amplified cDNA partial clone as probe. | | | | |
| The cDNA can be used to provide probes and primers useful e.g. in | | | | |
| detecting haemotogenous micrometastatic tumour cells and determining | | | | |
| prostate cancer progression (see also T36827-30 and T36813-18), | | | | |
| and in gene therapy. An alternatively spliced PSM, PSM' has a | | | | |
| shorter cDNA sequence. PSM genomic DNA is given in T36786. | | | | |
| Sequence 2654 BP; 782 A; 525 C; 639 G; 708 T; | | | | |
| atcaacaacatcttgcgtgcacatgaaggcgcgcacagacagatcaactagctgcatc 956 | | | | |
| 1324 ATTTCACAAAGTATAGTACGTCCTCAGAGGAGCAGTGGACACAGATATGTCAATTTCG 1383 | | | | |
| 957 ggggcccaaggagatgcacatggcccccaggagacagctaaatccgcgtgtggagacgtata 1016 | | | | |
| 1384 GGAGGTCACCGGGACTCATGGGTGTTTGGTGTGATTGACCTTCAGAGTGGACAGCTGTT 1443 | | | | |
| 1017 ctccctgagacgtgtgtcggacaccttctccatcgtgtgagcaacggcttccggcccgacaga 1076 | | | | |
| 1444 GTTCATGAAATTTGTGAGGAGCTTTTGACACATGAAAAAGAAAGGCGAGAGACTTGAAGA 1503 | | | | |
| 1077 agtcctcttcatacagcctggagacggtgtgtgaccttggagacgttggcccccagagtg 1136 | | | | |
| 1504 ACAATTTTGTGTCAGAGCGGATGACAGAAATTTGGCTTCTTGCTTACTGTAGTGG 1563 | | | | |
| 1137 ctgagaagcctacactagcgtgtgtgaccccaagccctgagtgtacgtgagcctgac 1193 | | | | |
| 1564 GCAGAGGAGAAATTCAGAGCTCCTTCAAGACGCGGGCGTGTATTATTAATGCTGAC 1620 | | | | |
| RESULT 7 | | | | |
| N70566 | | | | |
| ID N70566 standard; DNA: 3059 BP. | | | | |
| AC N70566; | | | | |
| DT 26-APR-1991 (first entry) | | | | |
| DE Sequence encoding functional part of corn phosphoenolpyruvate | | | | |
| DE carboxylase (PEPC ase). | | | | |
| KW Enzyme: CO2 fixation; corn starch; photosynthesis; ss. | | | | |
| OS Corn. | | | | |
| FH Key | | | | |
| FT cds | | | | |
| FT Location/Qualifiers | | | | |
| FT 3..2810 | | | | |
| FT /*tag= a | | | | |
| PN EP-212649-A. | | | | |
| PD 04-MAR-1987. | | | | |
| PF 22-AUG-1986; 111680. | | | | |
| PR 23-AUG-1985; JP-186181. | | | | |
| PA (SUMO) SUMITOMO CHEM IND KK. | | | | |
| PA (KATSU/) KATSUKI H. | | | | |
| PI Katsuki H. | | | | |
| DR WPI: 87-058191/09. | | | | |
| DR P-PSDB: P70362. | | | | |
| PT New dna sequence coding for corn | | | | |

| | Query Match | 1.9%; | Score 47.6; | DB 1; | Length 3059; |
|--------|---------------------------------------------------------------------------|-----------------|-----------------|-----------|--------------|
| | Best Local Similarity | 49.2%; | Pred. No. 0.24; | | |
| | Matches 187; | Conservative 0; | Mismatches 184; | Indels 9; | Gaps 2; |
| PT | phospho:enol-pyruvate:carboxylase - useful in recombinant dna | | | | |
| PT | procedures for expression of enzyme | | | | |
| PS | Claim 2; Fig 1; 27pp; English. | | | | |
| CC | PEPC ase plays an important role in photosynthetic CO2 fixation in | | | | |
| CC | corn or maize. Cloning of the gene encoding it, modification of the | | | | |
| CC | nucleotide sequence and integration of this modified gene into plant | | | | |
| CC | cultures can lead to increased fixation efficiency and good harvest | | | | |
| CC | of corn starch | | | | |
| SO | Sequence 3059 BP; 662 A; 934 C; 897 G; 566 T; | | | | |
| OY | 129 ctctcccgcttaggcaaccagcctcttggaacgggtgccaagctcgccggatgacc 188 | | | | |
| DB | 1239 CTCTCCCGGGGAACCTGGACATCCGGAGGAGTGGAGCGGCACACCGAGCTGATCGAC 1298 | | | | |
| OY | 189 gctctgactgaagcaattcgcgcgctctcccgccaagaagctgaccagttgtgagcc 248 | | | | |
| DB | 1299 GCCATCACACGACACCTCGCATCGGCTGTCATCCCGAGTGGTCCGAGACAAGAGCCAG 1358 | | | | |
| OY | 249 gacacgacctacgttggggtgcaattcccgagatccgagctacccccaacacctgtactg 308 | | | | |
| DB | 1359 GAGTGGCTGCTGTGGAGACTCGAGGCAAGGCCCGCGCTGTGCCCCCGACCTTCCAG 1418 | | | | |
| OY | 309 gtcgatgagcgccggaaggtcggaagagagctgcgcgtcggaagaccctgaagctacgac 368 | | | | |
| DB | 1419 ACCGACGAGATGCGCGAGTATCGGCGCGTTC-----ACGTCTTGGGAGGTCCCG 1472 | | | | |
| OY | 369 ccctcaagcgccatcggaacgtcaacggaagagctgtgtatcgcccaactagcgagcgcc 428 | | | | |
| DB | 1473 CCGGACACTTCGGGCCCTCATATCTTCATGAGGAGAGGCCCCCTCGAGCGTGTCCG 1532 | | | | |
| OY | 429 gaagacctcgcaagacctgcyggccagggcggtgatccagtggcgccgctctgtgtg 488 | | | | |
| DB | 1533 GTAGAGCTTCCTGCA---GCGGAGAGTGCGGCGTGGCGGACCGCTGCCGTGATCGCGCTG 1589 | | | | |
| OY | 489 cgcggtgggggtgacagctt 508 | | | | |
| DB | 1590 TTGAGAGCGTGGCCAGCCT 1609 | | | | |
| RESULT | 8 | | | | |
| V99230 | V99230 standard; DNA: 1030 BP. | | | | |
| AC | V99230: | | | | |
| DT | 09-MAR-1999 (first entry) | | | | |
| DE | DNA encoding an active acyltransferase domain (venat) from a PKS. | | | | |
| KW | Enzymatically active acyltransferase domain; polyketide synthase; PKS; | | | | |
| KW | polyketide synthetis; antimicrobial; anticancer; antifungal; | | | | |
| KW | immunosuppressant; antihelmintic; venat; malonate acyltransferase domain; | | | | |
| KW | PKS cluster; Streptomyces venezuelae ATCC15439; ds. | | | | |
| OS | Streptomyces venezuelae. | | | | |
| PN | W09851695-A2. | | | | |
| PD | 19-NOV-1998. | | | | |
| PF | 13-MAY-1998; U09518. | | | | |
| PR | 16-MAY-1997; US-858003. | | | | |
| PA | (ABBO) ABBOTT LAB. | | | | |
| PI | Kakavas SJ, Katz L, Pereda-Lopez A, Ruan X, Stassi DL, | | | | |
| PT | Summers RG; | | | | |
| PT | WPI; 99-070114/06. | | | | |
| DR | P-PSDB; W07713. | | | | |
| DR | New polyketides, produced by microorganism having altered polyketide | | | | |
| PT | synthase genes - especially chimeras containing a heterologous | | | | |
| PT | acyltransferase domain, useful as e.g. antimicrobials, anticancer | | | | |
| PT | agents | | | | |
| PS | Claim 24; Fig 13; 137p; English. | | | | |
| CC | The present sequence encodes an enzymatically active acyltransferase | | | | |
| CC | domain, venat, from a polyketide synthase (PKS). The venat protein | | | | |
| CC | is derived from the malonate acyltransferase domain from the | | | | |
| CC | PKS cluster of Streptomyces venezuelae ATCC15439. The sequence | | | | |

```

RESULT      9
ID          037948 standard; cDNA; 2353 BP.
AC          037948;
DT          11-JUL-1993 (first entry)
DE          Sequence of a DNA isolate encoding eukaryotic transcription factor
DE          (TF).
KW          Transcription factor; Y1; modulatory adeno-associated virus type 2;
KW          Epstein-Barr virus; oncogene; ss.
OS          Homo sapiens.
FH          Key
FT          poly_a_signal      Location/Qualifiers
FT          poly_a_signal      1494..1498
FT          poly_a_signal      /*tag= a
FT          poly_a_signal      2333..2338
FT          poly_a_signal      /*tag= b
FT          cds                241..1485
FT          cds                /*tag= c

MN          M0304076-A.
PD          04-MAR-1993.
PF          14-AUG-1992; U06840.
PR          16-AUG-1991; US-746485.
PA          (UYPR-) UNIV PRINCETON.
PI          Seto E, Shenk T, Shi Y;
PI          WPI; 93-093928/11.
DR          P-PSDB; R32020.
PT          Purified and isolated mammalian transcription factor Y1 - used
PT          to modulate transcription in adeno-associated virus type 2 -
PT          Epstein-Barr viruses and oncogenes; useful in diagnosis and
PT          screening
PS          Claim 32; Fig 11; 100pp; English.
CC          The cDNA of transcription factor (TF) Y1 (037948) was derived from
CC          HeLa cells derived from cervical carcinoma from clone p14-1 or pY1
CC          of the D98/AH-2 library. TF Y1 has a mol. wt. of 68 kD and is
CC          capable of binding to a sequence between -50 and -70 (PS-60 site)
CC          and to the transcription initiation region of the promoter (PS+1
CC          site) of an adeno-associated virus (AAV). It represents transcription
CC          directed by a RNA element plusinitiator (Inr) sequence. It can

```

| | | | |
|----|-----|----------------------------------------------------------------|-----|
| Oy | 39 | ccccccgccccccgccccctbgsagaccacagctctcgtatctctcccttccccc | 98 |
| Db | 96 | CCCCCGCCGCCGCCCTCTCTCCCCCGCCGCGCCCTTCTCTCTGTGCTTCT | 155 |
| Oy | 99 | tctctccgpgccccaanaaagctccaccctctcccgcttagagcaaacagccttcg | 158 |
| Db | 156 | TCCCGACGGCGCGGCTCTCTGCGCGCGCGCCGCGACGCGAGACCGAGGCGCGC | 215 |
| Oy | 159 | gaag-----gttgcaggtctcgcgcgagatgycgcctctactcaagacatlcgcgcgc | 214 |
| Db | 216 | GCCTGTGCGCGCGAGGCTCTAGCGCAATGGCTCGGCGACACCTCTACATCGCACGGA | 275 |
| Oy | 215 | gctctccgcagaaagctgtgcacacgtgttgcacgcagacacactacgttggcctgaatt | 274 |
| Db | 276 | CGGCTCGAGATGCCGCGCGAGATCGTGGAGCTCAGAGACTGAGAGTGAAGTGGAGACATCCC | 335 |
| Oy | 275 | cccgatctcgcctcaccccaacacctgtcaccttggctgatgagcgcggaagtlcgaga | 334 |
| Db | 336 | GGTGAGACCATTCAGACACACAGTGTGTGGCGAGAGAGAGAGACGACGACGACGA | 395 |
| Oy | 335 | gcacgtccgcttgaagaacctgtgcgtctactgtcccttagagagccatggcaagctcac | 394 |
| Db | 396 | GGACGCGCGCGGTGCGCACCGCGCGCGGGGGGCGCCACGGGACGCCCGCCACACCA | 455 |
| Oy | 395 | gggaagcttgttgcagcccaactacgcgcgcgcgcgaagacctctcaagacctctgcgcag | 454 |
| Db | 456 | CCACCACCATCACACGACGACGACCGCGCCCATGATGCTGTGTGAGCGCGTGGACACGA | 515 |
| Oy | 455 | gggcgttgatccagtgtggcgcgcctgcctgcctgtgtgcgtgtgggtgtatcagc | 506 |
| Db | 516 | CGACCCGACCCAGGTGCACGACGACGAGAGAGTATCTGTGTGAGACGCGCGC | 567 |

```

RESULT 10
V07530
ID V07530 standard; cDNA, 2353 BP.
AC V07530;
DT 09-NOV-1998 (first entry)
DE Human transcription factor YY1 cDNA.
KW YY1; transcription factor; human immunodeficiency virus; HIV; AIDS;
KW infection; retrovirus; therapy; LSF; ss.
OS Homo sapiens.
FH Key
FT CDS location/Qualifiers
241..1485
/*tag= a
PN W09833067-A1.
PD 30-JUL-1998.
PF 13-JAN-1998; U00574.
PR 23-JAN-1997; US-036242.
PA (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.
PI Devico A, Margolis D, Romero F;
DR WPI; 98-428099/36.
DR P-PSDB: W65406.
PI Transcription factors YY1 and LSF, and their derivative(s) and
PI analogue(s) - useful for, e.g. inhibiting HIV transcription.
PT replication and/or infection in vitro or in vivo or preventing
PT disorders associated with HIV; English.
PS Disclosure; Page 74-76; 112P; Enghish.
CC W65406), a multifunctional transcription factor that has previously

```


| | | | |
|----------------------------------------------------------------|-----------------------------------------------------------------------|-----------------------------------------------------------------|-------|
| OY | 1879 | gagccacacgcctggcgccccctgcgtggaccacctcgtcgctctgcttgcctccaacatccg | 1938 |
| | | | |
| Db | 28097 | TGACCTGCCCGAAGATCCCTTCACCAACCGGCCTACTGCTGTGGCCCGCCGGAATCCG | 28156 |
| OY | 1939 | ggaccgccgggac | 1951 |
| | | | |
| Db | 28157 | GCACGACGCGGC | 28169 |
| <hr/> | | | |
| RESULT | 12 | | |
| ID | Q37543/c | | |
| AC | Q37543 standard: cDNA; 4356 BP. | | |
| DT | 17-JUN-1993 (first entry) | | |
| DE | Cardiac adenylyl cyclase type V gene. | | |
| KW | CACV; therapy; diagnostic; cardiac function; cyclic AMP; cAMP; heart; | | |
| KW | failure; ss. | | |
| OS | Canis familiaris. | | |
| FH | Key | Location/Qualifiers | |
| FT | cds | 148..3702 | |
| FN | EP-529622-A. | /tag= a | |
| PD | 03-MAR-1993. | | |
| PF | 27-AUG-1992; 114637. | | |
| PR | 29-AUG-1991; US-751460. | | |
| PA | (AMCY) AMERICAN CYANAMID CO. | | |
| PI | Ishikawa Y, Komski AF; | | |
| DR | WPL; 93-06868/09. | | |
| P-PSDB: | R32882. | | |
| PT | Isolated nucleic acid mol. encoding Cardiac adenylyl cyclase type | | |
| PT | V - useful for determining and modifying cardiac function | | |
| PS | Claim 1; Page 15-27; 38pp English. | | |
| CC | Left ventricular tissue of canine heart was used as a source of mRNA. | | |
| CC | A cDNA library was prepd. in lambda gt10 phage. A 970 bp Aat-HincII | | |
| CC | fragment from type I adenylyl cyclase cDNA was used as probe. The | | |
| CC | clones isolated were used to obtain cDNA encoding CACV. This probe | | |
| CC | may also be used to screen a human cardiac cDNA library to obtain | | |
| CC | the cDNA encoding human CACV. CACV, its analogues and antibodies | | |
| CC | are useful in therapy or diagnostic assays, e.g. in modifying and | | |
| CC | determining cardiac function. A decrease in CACV content of the | | |
| CC | heart contributes to impaired CAMP prodn. and in heart failure. The | | |
| CC | CACV can also be used to screen for cpds. which stimulate or inhibit | | |
| CC | the activity of the cyclase. | | |
| SQ | Sequence 4356 BP; 837 A; 1367 C; 1377 G; 775 T; | | |
| <hr/> | | | |
| Query Match 1.8%; Score 45; DB 1; Length 4356; | | | |
| Best Local Similarity 49.4%; Pred.No. 0.86; | | | |
| Matches 117; Conservative 0; Mismatches 120; Indels 0; Gaps 0; | | | |
| OY | 1506 | gqccgcctgcgcgcgcgcgttgcccaaggcgcgtgcccagctgcgcgaagctcctaaccgg | 1565 |
| Db | 469 | GCCCTCGGGGCTTGGGCTTGCGCGCCGCCGCCGCCCTCGTGCACAGCTCCAC | 410 |
| OY | 1566 | ctcaagcacgatcgcgtctgcacctgcacttcggccgcgcgtacggggagctgtctcaag | 1625 |
| Db | 409 | CGAGCGGGGGCGCACACTCGCCCCCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCG | 350 |
| OY | 1626 | caaatcggaacctcaagaagtctcttgggagcctcaaggccgcggcgtgaccttcaa | 1685 |
| Db | 349 | GCGCGCCGGAGACTCGCCCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 290 |
| OY | 1686 | tgggtgctactcgcgccggggggagcatcataatccggcgggaaagtctcgcggaagag | 1742 |
| Db | 289 | CAGTCGTGCGCCGCGCGCTCTGCCAGCGCCGACCTGGAACGGAACTGAAGCGAAG | 233 |
| <hr/> | | | |
| RESULT | 13 | | |
| ID | Q95540/c | | |
| CD | Q95540 standard: DNA; 4356 BP. | | |
| DC | Q95540; | | |
| DT | 31-JAN-1996 (first entry) | | |

| | |
|----------------------------------------------------------------|---------------------------------------------------------------------------|
| DE | Cardiac adenyllyl cyclase gene. |
| KW | Cardiac adenyllyl cyclase; effector enzyme; ss. |
| OS | Homo sapiens. |
| FH | Key |
| FT | cds |
| FN | Location/Qualifiers |
| PD | TW-243453-A. |
| PF | 21-MAR-1995. |
| PR | 02-JUL-1992; 105242. |
| PA | 12-JUN-1992; US-899068. |
| DR | (AMCY) AMERICAN CYANAMID CO. |
| P | WPI: 95-214006/28. |
| PT | P-BSDb: R78519. |
| PS | Cardiac adenyllyl cyclase and corresp. DNA - having specified sequences |
| CC | Claim 1; Fig 2; 45pp; Chinese. |
| SQ | Sequence 4356 BP; 836 A; 1372 C; 1373 G; 775 T; |
| Query Match | |
| Best Local Similarity 49.4%; Pred. No. 0.86; | |
| Matches 117; Conservative 0; Mismatches 120; Indels 0; Gaps 0; | |
| Dy | 1506 ggcgcgctggcccgctggcccagcgttgcccagctgcaggcaactcctcatccgg 1565 |
| Db | 469 ggccctggcgctggcgctggcgccgccgccgccgccgctgctcagaccagctccac 410 |
| Oy | 1566 ctcagcacaagatcgctctgctgccttgaattcgccgctaagggagaagtgcctcagg 1625 |
| Db | 409 ccaggcggggcgccacctccgccccccccgccccccccgccggccggccggccgcmag 350 |
| Oy | 1626 cacatcggaacactaacagagttctcttgggagctccaaggcccgaggctgacctgcag 1685 |
| Db | 349 gccgccggggagactgccccccgccccccccgccccccccgccggctgctccggccccg 290 |
| Oy | 1686 tgggtgtactacgcgcgggggagtaatacacatccgcggcgagaaagtccggcaggag 1742 |
| Db | 289 cagtcgtgcgcccccgcgcctccgccaggccgacnctgagacgtagacccaag 233 |
| RESULT 14 | |
| V45824 | |
| ID | V45824 standard; DNA: 6781 BP. |
| AC | V45824: |
| DT | 09-NOV-1998 (first entry) |
| DE | Maihe phosphoenolpyruvate decarboxylase gene. |
| KW | Phosphoenolpyruvate carboxylase; PEPC; maize; corn; C3 plant; rice; |
| KW | C4 plant; photosynthesis; transgenic plant; ss. |
| OS | Zea mays. |
| FH | Key |
| FT | TATA_signal |
| FN | Location/Qualifiers |
| PD | 1185..1189 |
| FT | /*tag= a |
| FT | CAAT_signal |
| FT | 842..846 |
| FT | /*tag= b |
| FT | protein_bind |
| FT | 1128..1233 |
| FT | /*tag= C |
| FT | /note= "Sp1 binding site" |
| FT | 1160..1165 |
| FT | /*tag= d |
| FT | protein_bind |
| FT | 1488..1493 |
| FT | /note= "Sp1 binding site" |
| FT | /*tag= e |
| FT | protein_bind |
| FT | 1494..1499 |
| FT | /note= "Sp1 binding site" |
| FT | /*tag= f |
| FT | misc_feature |
| FT | 552..560 |
| FT | /*tag= g |
| FT | /function= light responsive element |
| FT | repeat_unit |
| FT | 663..680 |
| FT | /*tag= h |

```

FT      repeat_unit      /rpl_type= "DIRECT"
FT      686. .902
FT      /tag= i
FT      /rpl_type= DIRECT
FT      743. .762
FT      /tag= j
FT      /rpl_type= DIRECT
FT      818. .838
FT      /tag= k
FT      /rpl_type= DIRECT
FT      999. .1012
FT      /tag= l
FT      /rpl_type= DIRECT
FT      1174. .1183
FT      /tag= m
FT      /rpl_type= DIRECT
FT      1297. .6318
FT      /tag= n
FT      /note= "contains 9 introns"
FT      1297. .1476
FT      /tag= o
FT      /number= 1
FT      1477. .1586
FT      /tag= p
FT      /number= 1
FT      1587. .1981
FT      /tag= q
FT      /number= 2
FT      1982. .2091
FT      /tag= r
FT      /number= 3
FT      2177. .3073
FT      /tag= s
FT      /number= 3
FT      3074. .3296
FT      /tag= t
FT      /number= 4
FT      3297. .3745
FT      /tag= u
FT      /number= 4
FT      3746. .3849
FT      /tag= v
FT      /number= 5
FT      3850. .3974
FT      /tag= w
FT      /number= 5
FT      3975. .4065
FT      /tag= x
FT      /number= 6
FT      4066. .4202
FT      /tag= y
FT      /number= 6
FT      4203. .4357
FT      /tag= z
FT      /number= 7
FT      4358. .4453
FT      /tag= aa
FT      /number= 7
FT      4454. .5452
FT      /tag= ab
FT      /number= 8
FT      5453. .5549
FT      /tag= ac
FT      /number= 8
FT      5550. .5936
FT      /tag= ad
FT      /number= 9
FT      5937. .6024
FT      /tag= ae
FT      /number= 9

FT      repeat_unit      /rpl_type= "DIRECT"
FT      686. .902
FT      /tag= i
FT      /rpl_type= DIRECT
FT      743. .762
FT      /tag= j
FT      /rpl_type= DIRECT
FT      818. .838
FT      /tag= k
FT      /rpl_type= DIRECT
FT      999. .1012
FT      /tag= l
FT      /rpl_type= DIRECT
FT      1174. .1183
FT      /tag= m
FT      /rpl_type= DIRECT
FT      1297. .6318
FT      /tag= n
FT      /note= "contains 9 introns"
FT      1297. .1476
FT      /tag= o
FT      /number= 1
FT      1477. .1586
FT      /tag= p
FT      /number= 1
FT      1587. .1981
FT      /tag= q
FT      /number= 2
FT      1982. .2091
FT      /tag= r
FT      /number= 3
FT      2177. .3073
FT      /tag= s
FT      /number= 3
FT      3074. .3296
FT      /tag= t
FT      /number= 4
FT      3297. .3745
FT      /tag= u
FT      /number= 4
FT      3746. .3849
FT      /tag= v
FT      /number= 5
FT      3850. .3974
FT      /tag= w
FT      /number= 5
FT      3975. .4065
FT      /tag= x
FT      /number= 6
FT      4066. .4202
FT      /tag= y
FT      /number= 6
FT      4203. .4357
FT      /tag= z
FT      /number= 7
FT      4358. .4453
FT      /tag= aa
FT      /number= 7
FT      4454. .5452
FT      /tag= ab
FT      /number= 8
FT      5453. .5549
FT      /tag= ac
FT      /number= 8
FT      5550. .5936
FT      /tag= ad
FT      /number= 9
FT      5937. .6024
FT      /tag= ae
FT      /number= 9

FT      exon      6025. .6318
FT      /tag= af
FT      /number= 10
FT      6512. .6520
FT      /tag= ag
FT      6548. .6553
FT      /tag= ah
FT      AU-693754-B.
FT      PD 02-JUL-1998.
FT      PE 03-FEB-1998; 052880.
FT      PR 11-MAR-1997; JP-056742.
FT      PA (NORO ) NAT INST AGROBIOLOGICAL RESOURCES.
FT      PI Matsuoka M, Sun-Ben Ku M, Toki S, Tokutomi M,
FT      DR WPI, 98-428353/37.
FT      PT Transgenic C3 plant expressing C4 plant gene - coding for enzyme
FT      involved in photosynthetic pathway
FT      PS Example 1; Fig 2-4; 42pp; English.
FT      CC This is the maize phosphoenolpyruvate carboxylase (PEPC) structural
FT      gene. It was isolated from a maize etiolated leaf genomic DNA
FT      library by hybridisation to an oligonucleotide probe (see V45825).
FT      A C3 plant expressing a gene of a phylogenetically related C4 plant
FT      is new, where the C3 plant has DNA containing the expression
FT      control region of a gene for an enzyme involved in a photosynthetic
FT      pathway of the C4 plant and a structural gene for an enzyme involved
FT      in a photosynthetic pathway of the C4 plant, where the C3 plant
FT      expresses the enzyme encoded by the structural gene at a high level.
FT      Also claimed are a portion of a plant as above, and a C3 plant
FT      tissue expressing a gene as above. The invention is specifically
FT      described by exemplifying a maize PEPC genome gene as DNA
FT      containing an expression control region and a structural gene for
FT      an enzyme involved in a photosynthetic pathway of a C4 plant and
FT      rice as a C3 plant. Transgenic rice plants having a PEPC actively
FT      about 75 times as high as that of the non-transgenic parent have
FT      their CO2 compensation point decreased by about 10%. It is expected
FT      that the photosynthetic capacity of an agriculturally important C3
FT      plant will be markedly improved by providing the photosynthetic
FT      function of a C4 plant.
FT      SO Sequence 6781 BP; 1690 A; 1853 C; 1546 G; 1692 T;

Query Match 1.8%; Score 44.8; DB 1; Length 6781;
Best Local Similarity 47.1%; Pred. No. 1;
Matches 179; Conservative 0; Mismatches 192; Indels 9; Gaps 1;

QY 129 ctctcccgcttagcgaaacacgacctcggaacgggtgcaagctcggccggatgacc 188
DB 4562 CTCCTCCCTGCTGAGCTGAGCATCCGACGAGCTCGACGCGACACCGACGATCGAC 4621
QY 189 gctctgacctcaggaacttcggcggtgctctccgcgaagctggaacagtgtgacc 248
DB 4622 GCCATCACCACGACCTCGGCAATCGGGTGTACCGGAGTGTCCGAGGACAAAGCGGACG 4681
QY 249 gacacgactacgtgggtgctgcaattccggatcgggtccaccccaaacccctgacctg 308
DB 4682 GACTGCTGCTGTGCGAGCTGCGAGCGAAGCGCCCTGCTGCCCCGGACCTTCCACG 4741
QY 309 gtccatgagccggaaggtcggagagcagctgcgctgagagaccctgaccttactatgc 368
DB 4742 ACCGAGAGATCGCCGACGTCATCGCGCGCTTCC-----ACGTCTCGGGGAGCTC 4792
QY 369 cccctacagccatcgcaacgttcaagggagagctgtgttaagccactaaggcgagccc 428
DB 4793 CCGCCCGACAGCTTCGGCCCTTACATCATTCATGCGACGACGCCCTTCGGAGTGCTC 4852
QY 429 gaagacctcaggaaccttcgggccaaggtgagctgagcttggtggccgctgtctgtggtg 488
DB 4853 GCCGTGAGACTCTGTGACGGCGAGTGTGGCGTGTGCGCCACCGCTGCGCTGTGCGCTG 4912
QY 489 cgcgtgggggtgatacagctt 508
DB 4913 TTGGAAGGCTGGCCAGCCT 4932

```

RESULT 15

ID X26879 standard; DNA; 2135 BP.

AC X26879;

DT 23-JUN-1999 (first entry)

DE DNA encoding a protein with cation transporting activity.

KW Organic cation transporter; OC11; OC12; drug development; fatty liver;

KW heart disease; cancer; anti-tumour drug; anticancer drug; ss.

OS Homo sapiens.

FN key

FT CDS

FT Location/Qualifiers

FT 147..1802

FT /*tag= a

PN W09913072-A1.

PD 18-MAR-1999.

PE 07-SEP-1998. J04009.

PR 20-MAY-1998; JP-156660.

PR 08-SEP-1997; JP-260972.

PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

PI Mezu J, Oku A;

DR WPI: 99-215062/18.

DR P-PSDB: Y01649.

PT Genes homologous with organic cation transporters OC11 and OC12, useful in design of new drugs for treatment of diseases due to abnormality of the transporter functions

PS Claim 2; Page 45-51; 97pp; Japanese.

CC The present sequence encodes a protein with cation transporting activity. The genes are significantly homologous with organic cation transporters OC11 and OC12. The genes may used in drug development, particularly in the treatment of diseases due to abnormality of the organic cation transporter functions e.g. fatty liver, heart diseases and cancers, by controlling such as by inhibition or activation.

CC Administration of anti-tumour and anticancer drugs in combination with a transporter protein inhibiting agent allows the agents to penetrate into the diseased cells to enhance the drug action.

SO Sequence 2135 BP; 499 A; 530 G; 559 T;

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 13, 2000, 11:10:38 : Search time 1456.15 Seconds
(without alignments)
5929.991 Million cell updates/sec

Title: US-09-358-755-3

Perfect score: 2519

Sequence: 1 ggcgcgcggggagcgctct.....ctgatataaagcccaaa 2519

Scoring table: IDENTITY_NUC

Searched: 4089388 seqs, 1713955092 residues

Database: EST.*

Word size: 0

Number of hits that pass the threshold : 8178776

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: g_b_est1:*
21: g_b_est2:*
22: g_b_est3:*
23: g_b_est4:*
24: g_b_est5:*
25: g_b_est6:*
26: g_b_est7:*
27: g_b_est8:*
28: g_b_est9:*
29: g_b_est10:*
30: g_b_est11:*
31: g_b_est12:*
32: g_b_est13:*
33: g_b_est14:*
34: g_b_est15:*
35: g_b_est16:*
36: g_b_est17:*
37: g_b_est18:*
38: g_b_est19:*
39: g_b_est20:*
40: g_b_est21:*
41: g_b_est22:*
42: g_b_est23:*
43: g_b_est24:*
44: g_b_est25:*
45: g_b_est26:*
46: g_b_est27:*
47: g_b_est28:*
48: g_b_est29:*
49: g_b_est30:*

50: g_b_est31:*
51: g_b_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: g_b_est33:*
60: g_b_est34:*
61: g_b_est35:*
62: g_b_est36:*
63: g_b_est37:*
64: g_b_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: g_b_gss1:*
70: g_b_gss2:*
71: g_b_gss3:*
72: g_b_gss4:*
73: em_gss1:*
74: em_gss2:*
75: em_gss3:*
76: em_gss4:*
77: g_b_gss5:*
78: g_b_gss6:*
79: g_b_gss7:*
80: g_b_gss8:*
81: g_b_gss9:*
82: em_gss5:*
83: em_gss6:*
84: em_gss7:*
85: em_gss8:*
86: em_gss9:*
87: em_gss10:*
88: em_gss11:*
89: g_b_gss10:*
90: g_b_gss11:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 544.4 | 21.6 | 548 | 61 | AI814648 | AI814648 wj75f09.x |
| 2 | 521.6 | 20.7 | 536 | 38 | AA748421 | AA748421 ny01604.s |
| 3 | 481.4 | 19.1 | 700 | 50 | AJ225643 | AJ225643 RNJ225643 |
| 4 | 443.6 | 17.6 | 504 | 27 | AA007639 | AA007639 zh99608.r |
| 5 | 414.4 | 16.5 | 782 | 43 | AU035332 | AU035332 AU035332 |
| 6 | 394.8 | 15.7 | 452 | 23 | R99295 | R99295 yq71403.s1 |
| 7 | 389.6 | 15.5 | 444 | 22 | R46386 | R46386 yq50607.s1 |
| 8 | 383.6 | 15.2 | 425 | 22 | R49459 | R49459 yq58f02.s1 |
| 9 | 371.8 | 14.8 | 444 | 27 | AA007614 | AA007614 zh99608.s |
| 10 | 367.2 | 14.6 | 617 | 41 | AI047358 | AI047358 ud65602.y |
| 11 | 363 | 14.4 | 384 | 34 | AA505641 | AA505641 nh82603.s |
| 12 | 351 | 13.9 | 388 | 37 | AA682374 | AA682374 z186608.s |
| 13 | 350.2 | 13.9 | 363 | 38 | AA809686 | AA809686 n299604.s |
| 14 | 334.6 | 13.3 | 384 | 24 | H63101 | H63101 yf50h04.s1 |
| 15 | 334 | 13.3 | 366 | 22 | H70432 | H70432 yf50c12.s1 |
| 16 | 324.8 | 12.9 | 503 | 22 | R35943 | R35943 yq68f02.r1 |
| 17 | 318.2 | 12.6 | 341 | 37 | AA693893 | AA693893 z149407.s |
| 18 | 299.6 | 11.9 | 367 | 26 | W89178 | W89178 zh74h11.s1 |
| 19 | 292.6 | 11.6 | 708 | 41 | AI046365 | AI046365 ud65602.x |
| 20 | 267.4 | 10.6 | 408 | 21 | T64983 | T64983 yd11e08.s1 |
| 21 | 266.2 | 10.6 | 323 | 21 | T98167 | T98167 ye56c10.s1 |
| 22 | 236.2 | 9.4 | 421 | 45 | AI386238 | AI386238 mm20c12.y |

ALIGNMENTS

```
23 235.2 9.3 264 21 T98166 ye56c10.r1
24 225.8 9.0 465 21 T72291 yc68e08.r1
25 224 8.9 225 38 AA776476
26 220.6 8.8 384 48 A1551005
27 218.8 8.7 884 51 AU050369
28 218.6 8.7 1000 51 C81942
29 214.8 8.5 269 24 H63187
30 211 8.4 270 26 H63187
31 208.4 8.3 675 48 A1596094
32 203.2 8.1 655 44 A1256087
33 203 8.1 207 22 R29632
34 201.4 8.0 843 51 AU050545
35 199.2 7.9 436 34 AA51579
36 198 7.9 234 21 T87554
37 190.8 7.6 687 45 A1326398
38 183 7.3 461 43 A1195082
39 157.2 6.2 457 41 A1019669
40 134.6 5.3 458 64 A1120908
41 114.8 4.6 699 51 C81968
42 112.4 4.5 187 28 AA096819
43 101.4 4.0 106 24 H70431
44 100.8 4.0 462 81 A0673563
45 96.2 3.8 613 36 AA619986
```

```
RESULT 1
LOCUS A1814648 548 bp mRNA EST 24-AUG-1999
DEFINITION wj75f09.x1 NCI CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2408661 3'
similar to TR:O75422 OT5422 TRANSFERRIN-RECEPTOR2.1; mRNA
```

```
ACCESSION A1814648
NID 95425863
VERSION A1814648.1 GI:5425863
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
```

```
REFERENCE 1 (bases 1 to 548)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187613.
```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -400p from Gibco
High quality sequence stop: 460.
Location/Qualifiers
1. 548

FEATURES
SOURCE
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2408661"
/clone_lib="NCI-CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"
/dev_stage="adult"

/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

Query Match 21.6%; Score 544.4; DB 61; Length 548;
Best Local Similarity 99.5%; Pred. No. 5.3e-99;
Matches 545; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1968 ttccaggagagcgtttccggcgctagcagccctcagctgagcgtgagagagca 2027
Db ttccaggagagcgtttccggcgctagcagccctcagctgagcgtgagagagca 489
QY 2028 gccatgcgcttagcgggagatctcggaaacattgatacaactctgagccctggagat 2087
Db gccatgcgcttagcgggagatctcggaaacattgatacaactctgagccctggagat 429
QY 2088 cctcaatcccccgtcccccagtaagagctcctctgctcctcgttgatgaaggt 2147
Db cctcaatcccccgtcccccagtaagagctcctctgctcctcgttgatgaaggt 369
QY 2148 caggagagtggtgagagatgcacccctcattgctgataatttcattaccctcaaca 2207
Db caggagagtggtgagagatgcacccctcattgctgataatttcattaccctcaaca 309
QY 2208 tctctcaacgagaccagaccacacagatacacacacccagccctgagtgag 2267
Db tctctcaacgagaccagaccacacagatacacacacccagccctgagtgag 249
QY 2268 ctgaccttaattgagcgtgatactgtcggttaatacagaagtagcatccctcaatcac 2327
Db ctgaccttaattgagcgtgatactgtcggttaatacagaagtagcatccctcaatcac 189
QY 2328 agcccttcctccttctggtggtcctccatacctaagaagccatctggaggtttgctg 2387
Db agcccttcctccttctggtggtcctccatacctaagaagccatctggaggtttgctg 129
QY 2388 ccctggacctggccagcctctgtagtgaggagaagatcgtgacacatagccttaggcc 2447
Db ccctggacctggccagcctctgtagtgaggagaagatcgtgacacatagccttaggcc 69
QY 2448 aacaggtgctgtgtggaagggcggtgagttcaatatataaaccacgtgatatc 2507
Db aacaggtgctgtgtggaagggcggtgagttcaatatataaaccacgtgatatc 9
QY 2508 aataagc 2515
Db aataagc 1
```

RESULT 2
LOCUS AA748421/c
DEFINITION ny01a04.s1 NCI CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1270446 3'
similar to SW:TRSR_HUMAN P02786 TRANSFERRIN RECEPTOR PROTEIN ;
mRNA sequence.
ACCESSION AA748421
NID 92788379
VERSION AA748421.1 GI:2788379
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 536)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1798396.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.dio.llnl.gov/bdrip/image/image.html

Insert Length: 1516 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 385.

FEATURES
 source

1. 536
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="1"
 /clone="IMAGE:1270446"
 /clone_1db="NCI CGAP GCBI"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: p7713D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CEPR). cDNA synthesis was
 primed with a Not I - oligo(dt) primer
 [5'-TGTACCAATCTGAGTGGAGGCGGCTCATTTTCTTTTCTTTT-
 3']. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p7713 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 119 a 125 c 160 g 132 t
 ORIGIN

Query Match 20.7% Score 521.6; DB 38; Length 536;
 Best Local Similarity 99.2% Pred. No. 1.8e-94;
 Matches 524; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1992 cagctagccctgcactgcagctggaagggcagcgaatgcttagcgggagatgc 2051
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 536 CAGCTAGCCCTGCTOACTGAGCCTGCAAGGCGCAAGCAATGCGCTTAGCGGGAATGC 477
 QY 2052 tgaacaattgtaacaacttcgagccctgaggatcccaacccctcccaagca 2111
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 476 TGAACAATTGTAACAACACTTGTAGGCGCTTCTCATACATCCCGTCCCAAGCA 417
 QY 2112 agagctcctctgctcctcctgcttgatgattcaaggtcaaggaggtggtcagagtcacac 2171
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 416 AGAGCTCTCTGCTCCCTGCTTGAATGATTCAGGGTCAAGGAGGTGCTCAGAGTCCAC 357
 QY 2172 tctatgctgatatcaattctcattacccctcaacacatctcccaaggagcccaacccca 2231
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 356 TCTATGCTGATCAATTCATTAATCCCTCAACATCTCTCCACAGGAGCCCAAGCCCA 297
 QY 2232 gcacagatatcacacaccccaagccctgcaagttagctgaccccaatgtgaaggtcatcac 2291
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 296 GCACAGATATCTCACACACCCAGCCCTGCAATGATGACCTTAATGTGACGGTCAATAC 237

QY 2292 tctcgttaatcagagaagtatgacatcccttcattacacagcccttccttttgggtcc 2351
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 236 TCTCGTTAATCAAGAGATGACATCCCTTCATATACAGCCCTTCTTGGGGTCC 177
 QY 2352 lccataccagaagccactcctggaaggtttgctagagcccttggagccctggcagctcgtt 2411
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 176 TCCATACCTGAGACACACTCTGGGAGGTTGCTTAGGCCCTTGAGACCTGGCAGCTCTGTT 117
 QY 2412 agtggagagatgcgcctggcaccataagccttatggcacaaggtgtctgtgtgaaagg 2471
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 116 AGTGAGAGATGCTGCTGGACCATATAGCCTTATGCGCAACAGGTGTGTGTGTGAAGGG 57
 QY 2472 gctgtgaggttccatcatcaataaacccctgatatcaataagcccaaa 2519
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 56 GCGTGGAGTTTCATATCAATTAACACCATGATATCAATATAAAAAA 9

RESULT 3
 AJ225643 700 bp mRNA EST 14-MAY-1999
 LOCUS RNU225643 Rat liver ESTs (E.Olivier) Rattus norvegicus cDNA clone
 DEFINITION IRL413, mRNA sequence.
 ACCESSION AJ225643
 NID 94833473
 VERSION AJ225643.2 GI:4833473
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 700)
 AUTHORS Olivier,E., Soury,E., Risle,J.L., Smith,F., Schneider,K.,
 Lochner,K., Jouzeau,J.Y., Fey,G., and Sallier,J.P.
 TITLE A novel set of hepatic mRNAs preferentially expressed during an
 acute inflammation in rat represents mostly intracellular proteins
 JOURNAL Genomics 57 (3), 352-364 (1999)
 MEDLINE 99263497
 COMMENT On May 6, 1998 this sequence version replaced gi:3114939.

Contact: E. Olivier
 U78 INSERM
 543 chemin de la Breteque, 76233 Bois-Guillaume, France
 Library construction: J.B.C. Vol.270, p29998-30006 (1995)
 POLY-A-No.

FEATURES
 source

1. 700
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="IRL413"
 /clone_1db="Rat liver ESTs (E.Olivier)"
 /tissue_type="liver"
 /dev_stage="adult"
 /note="Organ: liver; Vector: lambda ZAPII; Library
 construction: Ripberger,J.A. et al. J. Biol. Chem. (1995)
 270(50):29998-30006"

BASE COUNT 149 a 198 c 210 g 143 t
 ORIGIN

Query Match 19.1% Score 481.4; DB 50; Length 700;
 Best Local Similarity 84.8% Pred. No. 2e-86;
 Matches 598; Conservative 0; Mismatches 101; Indels 6; Gaps 5;

QY 1106 tgacttggagcgtggtgctccagaggtggtcgaagcctcactgagcgtgctgacct 1165
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 TGACTTTGGAGCGGTGGAGCCACGGAGTTG-TGGAGGCTCACTCAGCGTCACTACT 59
 QY 1166 caaagccgtagtgaactgagcctggaacagcagctggtggaggtatgacaagtctcatgc 1225
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 60 CAAAGCGTGTGTATGTAGAGCCGTGACAACTCGTGTGGGAAACGGCAATTCATTCG 119
 QY 1226 caagacagccccccttgacaagctcattagagaagtgtcctgaagcaggtgtgattctcc 1285

| | | | |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|-------------|
| Db | 339 | CCAGGATCCCAACAAGCCGCTGTCTAGGCCACACAGGCTGTGTACGACATGTGTACACT | 398 |
| Qy | 644 | gggaactggaagacccctacaacccctggtctcccttcataatcaaacaggttcctcc | 703 |
| Db | 339 | GGGAACCTGGAGACCCCTTACACACCTCGCTTCCCGCTCTTCATCAACACCGATGCCCTTC | 458 |
| Qy | 704 | agttcatalcatcaaggcttccagatccacagccacccatcagttgacagatgtctc | 763 |
| Db | 459 | AGTAGAATCATCAGGCGCTTCCAGCATCCCGCGCCAGCCCATCATGTGCTGACATGTCTGA | 518 |
| Qy | 764 | ccgcttctgtaggaagctcaaggccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt | 823 |
| Db | 519 | TCAATGTGCTCAGGAACCTACACAGGCGCCGCTGTGCTTCCAGAGANTGGAAAGGACCTMTN | 578 |
| Qy | 824 | agggccccccttcaacccggggcccccggggccacagctgagtgagtgagtgagtgagtgag | 883 |
| Db | 579 | AGGCTCTCTCTTATTTGTGCTGGGACCTGGGCGCCGACCTTACGCTTGGAGGCGCAACACCG | 638 |
| Qy | 884 | ga--ccctccaccccatcaacaacacatcttcggtcgtcgtcgtcgtcgtcgtcgtcgtcgtc | 927 |
| Db | 638 | GAGCTCTTNTACCCCATATAGTAACATTTTGGGTGTCATGTGANGGC | 684 |
| RESULT | 6 | | |
| R99295/c | | | |
| LOCUS | R99295 | 452 bp | EST |
| DEFINITION | yg71d03.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone | | 14-SEP-1995 |
| ACCESSION | R99295 | | |
| VERSION | g985896 | | |
| KEYWORDS | R99295.1 GI:985896 | | |
| SOURCE | EST. | | |
| ORGANISM | human. | | |
| REFERENCE | Homo sapiens | | |
| AUTHORS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| | 1 (bases 1 to 452) | | |
| | Hillier,L., Clark,N., Dubque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R. | | |
| TITLE | The Mashu-Merck EST Project | | |
| JOURNAL | Unpublished (1995) | | |
| COMMENT | On Nov 29, 1993 this sequence version replaced gi:430583. | | |
| | Contact: Wilson RK | | |
| | Washington Unioersity School of Medicine | | |
| | 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 | | |
| | Tel: 314 286 1800 | | |
| | Fax: 314 286 1810 | | |
| | Email: est@wustl.edu | | |
| | Insert Size: 1674 | | |
| | High quality sequence stops: 262 | | |
| | Source: IMAGE Consortium, LNL | | |
| | This clone is available royalty-free through LNL; contact the | | |
| | IMAGE Consortium (info@image.lnl.gov) for further information. | | |
| | Insert Length: 1674 Std Error: 0.00 | | |
| | Seq primer: Promega -2im13 | | |
| | High quality sequence stop: 262. | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..452 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="GDB:3770271" | | |
| | /db_xref="taxon:9606" | | |
| | /map="21" | | |
| | /clone="IMAGE:201221" | | |
| | /clone_lib="Soares fetal liver spleen INFLS" | | |
| | /sex="male" | | |
| | /dev_stage="20 week-post conception fetus" | | |
| | /lab_host="DH10B (ampicillin resistant)" | | |
| | /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) | | |

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1734
High quality sequence stops: 326 Source: IMAGE Consortium, LNLN
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.lln.gov) for further information.
Insert Length: 1734 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 326

| | Query Match | Similarity | 15.2% | Score 383.6 | DB 22 | Length 425 |
|----|-------------|----------------------------------------------------------------|--------------|-------------------|---------------|------------|
| | Best Local | Similarity | 96.9% | Pred. No. 5.2e-67 | | |
| | Matches | 412 | Conservative | 0 | Mismatches 10 | Indels 3 |
| | | | | | | Gaps 2 |
| Qy | 2097 | ccgcgtccccaatcaagaagct-cctctgtctctcgttgaa--tgattcagggtcaggga | 2153 | | | |
| | | | | | | |
| Db | 425 | ccggtccccaatcaagagctccctctgtctcctcgtttaatgattttagggtcaggga | 366 | | | |
| Qy | 2154 | ggttgctcagagtcacacctcatgtgtgatcaatttcattacccctacatcttc | 2213 | | | |
| | | | | | | |
| Db | 365 | ggttgctcagagtcacacctcatgtgtgatcaatttctattacccctacatcttc | 306 | | | |
| Qy | 2214 | cacgagagcccaagaccccaagacagatatccacaaccccaagccctgcagtgtagctgacc | 2273 | | | |
| | | | | | | |
| Db | 305 | cacgagagcccaagaccccaagacagatatccacaaccccaagccctgcagtgtagctgacc | 246 | | | |

| RESULT | 9 |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| LOCUS | AA007614/c |
| DEFINITION | AA007614 444 bp mRNA EST 09-MAY-1997 z199608.s1 Soares fetal liver spleen JNLS.S1 Homo sapiens cDNA clone IMAGE:429423 3' similar to PIR-A8592 A8592 transferrin receptor protein - Chinese hamster ;, mRNA sequence. |

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel.: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL : contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 652 Std Error: 0.00
 Seq primer: mob.REGA+ET
 High quality sequence stop: 337.
 Location/Qualifiers
 1..444
 /organism="Homo sapiens"
 /db_xref="GDB:1329192"
 /db_xref="taxon:9606"
 /clone="IMAGE:429423"
 /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT773D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI
 This is a subtracted version of the Original Soares fetal

| Db | 315 | GCCTTTGCGGGGAGCCACGCTGTGGATGTTCTCTTCATGAGAGATGATGGGTGACCA | 374 |
|------------|----------------------------------------------------------------|------------------------------------------------------------------|------|
| Qy | 1455 | ttctgtacacaaagaggagacacttatatggaacctcacaagatgctgcaagccgctg | 1514 |
| Db | 375 | TTCTGTGACACGAAGAGGAGACACATATGAAATCTTCACAAAGATGCTGGAGGTGGCTG | 434 |
| Qy | 1515 | ccgcgcgttgccccagagccgtgagccacgtcgacagggacagtcctcatccgcgtcagcac | 1574 |
| Db | 435 | CCGCGCTGGGTGACGAGCGCATGTGGCTCAGCTCGCGGCGGCGACCTCTCATCCGATGAGCCAC | 494 |
| Qy | 1575 | gatgcctgtgctgccccctcgaccttcgagccgtacagggagacgtcgctccacagacatcg | 1634 |
| Db | 495 | GATGACTACTGTGCGGTACACTTCGTGGCGGCTATGAGAGAGTGGTCTTCAGGCAACATCGGC | 554 |
| Qy | 1635 | aacctcaacagatctctctggagaccctcaagccccgggagctacacctgaagtgggtgac | 1694 |
| Db | 555 | AACCTCAATAGATTCTCTGTGGGACACCTCAAGAGACGCGGCGTACCCCTGCATGGGTGTAC | 614 |
| Qy | 1695 | tc 1696 | |
| Db | 615 | TC 616 | |
| RESULT 11 | | | |
| AA505641/c | | | |
| LOCUS | AA505641 | 384 bp | mrna |
| DEFINITION | m182a03.s1 NCI_CGAP.Br1.1 Homo sapiens CDNA clone IMAGE:964972 | 18-AUG-1997 | |
| ACCESSION | AA505641 | | |
| NID | 92241778 | | |
| VERSION | AA505641.1 | GI:2241778 | |
| KEYWORDS | EST. | | |
| SOURCE | human. | | |

| | |
|-----------|--------------------------------------------------------------------------------------------------|
| REFERENCE | 1 (bases 1 to 384) |
| AUTHORS | NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap . |
| TITLE | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), |
| JOURNAL | Tumor Gene Index |
| COMMENT | Unpublished (1997) |
| | On May 9, 1995 this sequence version replaced gi:802599. |

Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LEML at:
www-bio.llnl.gov/bdrrp/image/image.html

Insert Length: 2525 Std Error: 0.00
Seq primer: -40m13 fwd ER from Amersham
High quality sequence stop: 256.
Location/Qualifiers
1..384

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:964972"
/clone_lib="NCI-GAP_Brl.1"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
/note="vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from pooled bulk
breast tumor tissue, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pRT3

```

```
vector. Library is not normalized. (The normalized
version of this library is NCLCGNP_Br2.) Library was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      85 a      86 c      108 g      105 t
ORIGIN
```

| | | | | |
|-----------------------|----------------|--------------|----------|------------|
| Query Match | 14.4% | Score 363 | DB 34 | Length 384 |
| Best Local Similarity | 98.7% | Pred. No. 6 | 4e-63 | |
| Matches 366 | Conservative 0 | Mismatches 5 | Indels 0 | Gaps 0 |

QY 2149 agggaggtgctcagaggtccacctctcatlgtgatcaatttctcatlaccctacacat 2208

Db 384 AGGAGGTGGCTCAGAGTCCACCTCTCATTTGCTGATCAATTTCTCATTTACCCCTACACAT 325

```

22  z205 cccccccagcccccaagacacagacacccccccagccccagcagc  z206
    |||||

```

[illegible]

Db 264 TGACCTAATGTGACGGTCTACTGTCGGTTAATCAGAGAGTAGCATCCCTTCAATCACA 205

QY 2329 qccccccttctctqqaqtcctccatacctaqaqaaccactctqqaqatttqctaqc 2388

Db 204 GCCCCTTCCCTTCTGGGGTCTCCATACCTAGAGACCACCTCTGGAGTTGCTAGGC 145

QY 2389 cctggacctggccagctctgttagtggagagatcgctggcaccatagccttatggcca 2448

Db 144 CCTGGGACCTGGCCAGCTCTGTAGTGGGAGAGATGGCTGGCACCATTAGCCTTATGGCCA 85

2449 acaggctggctctgtcggcgaaggggcgctggagtttcaatcattcaataaacaccctgtattca 2508

2 0 4 2

[illegible]

RESULT 12
AA682374/C

| DEFINITION | zj86a08.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens CDNA |
|------------|-----------------------------------------------------------------|
| LOCUS | 200 bp |
| FEATURES | 1..200 |
| ORIGIN | 1..200 |

PROTEIN (HUMAN) ; , mRNA sequence.
 22682374

NID 92669655
VERSION AA682374.1 GT:2669655

| KEYWORDS | EST. |
|----------|--------|
| SOURCE | human. |

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

| REFERENCE | 1 (bases 1 to 388) |
|---------------------|-----------------------------------------|
| HILLIER L. ALLEN M. | BOWLES L. DUBUCQUE T. CAISEL C. JOST S. |

Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,

WashU-NCI human EST Project

COMMENT On Sep 12, 1996 this sequence version replaced gi:1394174.

Contact: Wilson RK
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Email: est@watson.wustl.edu

IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primers: -40m13 fwd ET from Amersham

High quality sequence stop: 380.

FEATURES

SOURCE

Location/Qualifiers
1. 388
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="8"
/clone="IMAGE:461750"
/clone_1lb="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p7T73D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen 1NFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', ACTGGAAGATTATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p7T73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 65 a 108 c 129 g 86 t

ORIGIN

Query Match 13.9%; Score 351; DB 37; Length 388;
Best Local Similarity 97.3%; Pred. No. 1.6e-60;
Matches 357; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

379 ccatacgcaacgtcacgagagctgtgtacgcccactacagcgccgcaagacgtgc 438
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 388 CCATGGCAGACGTACGGGAGAGCTGTGTAGCCCCACTTACGGGGGCCGGAAGACCTGC 329

439 aggacctcgaggccagggcggtgatccagtggcgccgtctgtgtgtgtgtgtgt 498
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 328 AGGACCTCGGGGCCAGGGGGCTGTGATCCAGTGGGCCCTCTGTGTGTGCGGGGGG 269

499 tgatagctgtcccgagaaggtgacaaatgtctcaagacttgggggtcaagaagtgtcca 558
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 268 TGATCAGCTTTGGCCCGAAGAGTGAACCAATGCTCAGCACTTGGGGCTTAAGAGAGTGTCTCA 209

559 tataccagagccagcgagacttctccagagaccacccaagcctgtgccagccagc 618
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 208 TATACCCAGAGCGACGAGCTTCTCCAGGAGCCACCAAGCAAGCCTGTGCCAGCCAGC 149

619 aggcagctataggacatgtgcaactgggaacttgagaccctcaacactgtgtccctt 678
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 148 AGGCGGTATGAGCATGTGACCTGGGAGACTGGAGACCCCTACACACTGGCTTCCCTT 89

679 ccttcaatcaaacccagttccctcctcctgttgcatactcagagccttccagagcctccagagcc 738
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 88 CTTTCAATCAAAACCAGTTCCCTCCAGTTGATATCAAGGCCCTTCCAGCATCCACGCC 29

739 agcccat 745
|||
Db 28 GAGCCCT 22

RESULT 13
AA809686 363 bp mRNA EST 18-FEB-1998
LOCUS AA809686/c n299d04.s1 NCI-CCGAP_GCB1 Homo sapiens cDNA clone IMAGE:1303591 3',
DEFINITION mRNA sequence.
ACCESSION AA809686
NID 92879092
VERSION AA809686.1 GI:2879092
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 363)
AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2286784.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at:
www-bio.llnl.gov/bdrip/image/image.html

Insert Length: 1482 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 361.
Location/Qualifiers

FEATURES

SOURCE

1. 363
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="954B04: 954B04"
/clone="IMAGE:103591"
/clone_1lb="NCI-CCGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: p7T73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-GTGTACCAATCGAAGTGGAGCGGCCGCCGTCATTTTCTTTTCTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 83 a 77 c 106 g 97 t

ORIGIN

Query Match 13.9%; Score 350.2; DB 38; Length 363;
Best Local Similarity 99.2%; Pred. No. 2.2e-60;
Matches 352; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

2165 gtccacctcctcatgtcgtatcatcttctcattaccctacacatctctcagagagccca 2224
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 363 GTCCACCTCTCATGTGATCAATTTCTCATTCCTTACACATCTCTCCACGAGCCCA 304

2225 gaccccagcacagatcacacacacccccagccctgcagctgtatgtgacctaatgttagc 2284
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 303 GACCCAGACAGATATCCACACACCCACCCCTGCAGTGTGAGCTGAAGTGACG 244

2285 gtcaactgtcgtgttaacagagagtagcatcccttcaatcaacagcccttcccttctc 2344
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 243 GTCACTACTGTGCTTAATCAGAGAGTAGCATCCTTTAATCAGACGCCCTTCCCTTTCT 184

2345 ggggtctccatattcagagaccactctgggaggtgtgtctagagcctggagacctggccag 2404
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 183 GGGGTCTCTCAATACCTAGACACCACTCTGGAGAGTTGTCTAGGCCCTGGAGACCTGGCCAG 124

2405 ctctgttagtgggagagatgctgtgacacccatagccttaatggccaacgggtgtgtgtgtc 2464
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 123 CTCGTGTAGTGGGAGATGCTGGACACATAGCCTTAATGGCCAAAGGTGTGTGTGTGT 64

2465 gaaaggcgctggaggttcaataatcaataaacaccgagatataagccaana 2519
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

| | | | |
|------------|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|
| Db | 324 | CNACTACGGGNGGCCCGGAACACCTGCAGSAGCACTGGGGCCAGGGGCGGTGATCCAGTGGC | 265 |
| OY | 473 | ccgcgctcgtcgttgtagcgctggggggtg-atcagcttcgcccaaggtagcaatgctc | 531 |
| Db | 264 | CCGGCTGCTCTGTTGTCGGCGTGGGGGTGAATACANATTGACCAGGAGGTGACCATGCTC | 205 |
| OY | 532 | aggacttcgggggtccaaggagtgcataaccagaagcagcgactctccaggacc | 591 |
| Db | 204 | AGGACTTCGGGGGCTCAAGAGATGCTCATATPACCCAGAGACGACGAGACTTCTCCAGGACC | 145 |
| OY | 592 | caccacaagccaagccctgtccagccagcagcgagtataggaatatgtcaccttggaaactg | 651 |
| Db | 144 | CACCCAAGCCAAACCCGTCCAGCACACAGCAGTAGTATGGAATATGNCCTCGGCAATG | 85 |
| OY | 652 | gagacccctacaacctgtgcttccttccttaaccaaacccagttccctccagtgcatt | 711 |
| Db | 84 | GAGAACCCCTACACACTGGCTTCCTCTTAATCAACCAACCAAGTTCCTCCAGTTGCAT | 25 |
| OY | 712 | catcagagcttcccagcatccca | 734 |
| Db | 24 | CATCAGGCGCTTCCCAGCATCCCA | 2 |
| RESULT | 15 | H70432 | 366 bp mRNA EST 24-Oct-1995 |
| H70432/C | | YSO312.s1 Soares fetal liver spleen INFLS Homo sapiens cdna clone | |
| LOCUS | | IMAG3E:213718 3', mRNA sequence. | |
| DEFINITION | | H70432 | |
| ACCESSION | | H70432 | |
| NID | | 91040638 | |
| VERSION | | H70432..1 | |
| KEYWORDS | | EST. | |
| SOURCE | | human. | |
| ORGANISM | | Homo sapiens | |
| | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | |
| | | Eutheria; Primates; Catarrhini; Homnidae; Homo. | |
| REFERENCE | | 1 (bases 1 to 366) | |
| AUTHORS | | Hillier,L., Lennon,G., Becker,M., Donald,M.F., Chiapelli,B., Chisoe,S., Dietrich,N., Dubugue,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Trevisan,R., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasakis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M. | |
| TITLE | | Generation and analysis of 280,000 human expressed sequence tags | |
| JOURNAL | | Genome Res. 6 (9), 807-828 (1996) | |
| MEDLINE | | 97044478 | |
| COMMENT | | On Apr 14, 1993 this sequence version replaced gi:837502. | |
| | | Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: estelw@wustl.edu Insert Size: 801 High quality sequence strops: 269 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 801 Std Error: 0.00 Seq primer: Promega -21ml3 High quality sequence stop: 269. Location/Qualifiers 1 .366 /organism="Homo sapiens" /db_xref="GBD:3779022" /db_xref="taxon:9606" /map="11" /clone="IMAGE:213718" /clone_lib="Soares fetal liver spleen INFLS" /sex="male" | |
| FEATURES | | | |
| source | | | |

